

09/463705

428 Rec'd PCT/PTO 25 JAN 2000

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: Hoechst Marion Roussel
- (B) STREET: 1, Terrasse Bellini
- (C) CITY: PUTEAUX
- (E) COUNTRY: FRANCE
- (F) POSTAL CODE (ZIP): 92800
- (G) TELEPHONE: 01.49.91.57.27
- (H) TELEFAX: 01.49.91.46.10

- (ii) TITLE OF INVENTION: Biosynthesis and transfer genes of
6-deoxyhexoses in Saccharopolyspora erythraea and in
Streptomyces antibioticus and their use.

- (iii) NUMBER OF SEQUENCES: 61

(iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(vi) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: FR 9709458
- (B) FILING DATE: 25-JUL-1997

(vi) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: FR 9807411
- (B) FILING DATE: 12-JUN-1998

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3439 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Saccharopolyspora erythraea*

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION:complement (48..1046)

(D) OTHER INFORMATION:/function= "involved in the
biosynthesis of mycarose"
/gene= "eryBII"

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION:complement (2322..3404)

(D) OTHER INFORMATION:/function= "involved in the
biosynthesis of desosamine"
/gene= "eryCII"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GCTTCACGCT CACCAGCCGT ATCCTTTCTC GGTTCCTCTT GTGCTCACTG CAACCAGGCT	60
TCCGGCGCCG CGCCGCCGGA GGCCACCGCG GGGAAGATCT CGTCCAGTTC GGACAGCGCC	120
TGCTCGTCCA GGGTCATCGC GGACGCCTTC AGCGCGGAGT CGAGCTGCTC GGGGGTTCGC	180
GGGCCGATGA CGGCGCCGGC GATGCCGGGC CGGGACAGCA CCCATGCGAG CCCCACCTCG	240
GCCGGGTCTT CGCCGAGGTT GCGGCAGAAC TTCTCGTAGG CCTCGATCGC CGGGCGCAGG	300
GACGGCAACA GCACCTGCGC ACGGCCCTGC GCCGACTTCA CCGCGGTGCC CGCGGCCAGC	360
TTCTCCAGCG CTCCGCTGAG CAGGCCGCCG TGCAGCGGCG ACCAGGCGAA GACGCCGAGC	420
CCGTAGGCCT GCGCGGCGGG CAGCACCTCC AGTCGGCGT GCCGGACCGC CAGGTTGTAC	480
AGGCACTGGT GGGAGACCAT GCCCAGGGAG TGGCGGCGGG CGGCGTTCTC CTGCGCGGCG	540
GCGATGTGCC AGCCCGCGAA GTTCGACGAG CCGACGTAGG AGACCTTGCC GCTGGCGACG	600

AGGCTGTCCA TGGCCTGCCA CACCTCGTCC CACGGCGCGG ACCGGTCGAT GTGGTGCATC 660
TGGTAGACGT CGATGTGGTC GACGCCAGC CTGCGCAGCG ATCCCTCGCA GGAGGCGATG 720
ATGTGCCGCG CCGACAGCCC GCTGTCTGTTG ACGCGCTCGC TCATCTCGCC GCCGACCTTG 780
GTCGCCAGCA CGGTGTCCTC GCGCCGTCCG CCGCCCTGGG CCAGCCACCT GCCCACCAGC 840
TCCTCGGTGT GGCCCTTGTA GAGCCGCCAG CCGTACATGT CGGCGGTGTC GAGGCAGTTG 900
ATGCCGCGGT CCCGGGCGTG GTCCATCAGG CGCAGCGCGT CGTCGTCCTC GACGCGTCCG 960
CTGAAGTTCA CCGTGCCGAG CCAGAGCCTG CTGGTGAGCA GCGCGGAACG CCCGAGCCGC 1020
ACGTGCGTCG CGGCGTCGGT GGTTCATCGTG GTTCTCTCCT TCCTGCGGCC AGTTCCTCGC 1080
AGATGCCGAC GACCTCGGCC GGTGACGGCT CCGCGAGCAT GTCGTCGCGC ATCCGCGCCG 1140
CGCCGGCGCG GTGGGCGGGG TCGTCGAGGA CCCGCTTAC CCACTCCCGG AGCTGGTCGG 1200
GGGTCAGCTC GGGCACGGGC AGCGCGATCC CCGCCCCGAA TTCCTGCGTG CGCTGCGCGC 1260
GCACGCCGGT GTCCAGCCG TCGGGCAGGA TCACCTGCGG CACGCCGTGG ATCGCCGCGG 1320
TGTGCCAGCT CCCGGGTCCG CCGTGGTGCA CCGTCGCCG GCAGGTCGGC AGCAGCGCGT 1380
GCATCGGGAC GAAGCCGACC GTGCGGACGT TGTCCGGGAT GTTCGCGACG CTTTCTAGCT 1440
GCTGCGCGTC GAAGGTCGCG ATGATCTCGG CGTCGACGTC GCCGACGGCA CCCAGCAGCT 1500
CCTCGATGGA GACCTGCCCC ATGCTGTTCT CGCGGCTGGA GATCCCGAGC GTGAGGCACA 1560
CGCGGCGGCG CTCGGGCTCG TCGTGCAGCC ATTCCGGCAC CACGGACGGC CCGTTGTAGT 1620
CGACGTAGCG CATCCGACG GTCTTCAGGC CCGTGTCGAG CCTGATCGCG GCCGGGGCGG 1680
GGTCGATCGT CCACTGCCCC ACGACCACCT CCTCGTCGAA GGCCGGGCGG CCGTACTTCT 1740
CCAGCGTCCA GGTGAGCCAC TCGGCGAGCG GGTCTCTCCG GTGCTCCTCC GGCTGGTCGG 1800
GCAGCAGGCC GAGGAAGTTC TGCCGCGCCC GGGTGGTGAT GTCGGGTCCC CACAGCAGCC 1860

GCGCGTGCGG CGTTCGGGTC ACCGCCGCCG CGATGGGCGC GGCGAAGGTG AGCGGCTCCC 1920

AGATGACCAG GTCGGGCCGC CACTTCCGGC AGAACGAGAC CATGCCTTCG ATGAGCGTGT 1980

CCGGGCTCAT CAGGGCGTAG AAGGTCGGGG TGAGCACGGT CTGCATGCCC AGCAGGTGCT 2040

CCCAGGTCAA GGTGGCGGGG TCCCCTCGC TGAAGTCCAG GCTCCGGACG TAGTCGATGA 2100

TGTCGTGGCC CGCGTGGGTC ATGAAGTCCA CGAGGTCGAC GTCGGTGCCG ACCGGGACGG 2160

CGGTCAGCCC GGCCGCGGTG ATGTCCTCGG TGAGCGCCGG GGACGCGACC ACGCGGACCT 2220

CGTGCCCCGC CGCGCGGAAC GCCCATGCGA GGGGGACGAG GCCGAAGAGG TGGCTCTTGC 2280

TGGCCATGGA GGAGAAGACG ACGCGCATCG CGGTTACCTC AGAGCTCGAC GGGGCAGCGG 2340

TTGGTTCCCC GCAGGACGGG TGATCGGCGG CGCCGGACGA CCGGGCCGCT GGGCGTGAGT 2400

CCGGGCAGCG CCTTGGCCGC GGCCCGCAGT GCGGCGGTGG CGAGCGCGGT GACCAGCTCC 2460

TCCAGCTGC CGGGGTGGCC GCGATGTGCC GACAGCGCGC GGTCGGCGTC GGGGCGGTCC 2520

ACGTCGAGGC GGTGCGGGTC GGCGAAGACC TCCGGGTCGC GGTGGCCGC CGCGACGACG 2580

ACCACGACCT CCTCGCCTTC GCCGATCAG TGCTCGCCGA GCCGCACCTC TCGGTGGCC 2640

GTGCGCCGCT CCAGGTGCAA TGCCGGGTGC AGGCGCAGCA CCTCGGCGAC GGTTCGCTGC 2700

GCGGCGGCGG GGTGTCGGC GATCCGTTTCGCCAGCCCCG GTTCGGCCGA GACGGCCAGG 2760

ACCGCGTCGA CCACGGTGTT CGCGGTCATC TCGGCCCCGG CGAACAGGGC GCGCAGTGCG 2820

GGGTCGGCGG GCAGTGCCGC GACCGCTGCT TCGGTCACCG CGAGTGCTG CGGGCTGAGC 2880

TGGGCGTCCA GGCTGACGCG GCGTCCCAC GCGGCGCCGC GCAGCACTCC GGCTGCGCCG 2940

AGCACGGCGG TCATGCCCTG CACCGGTACC TGCCAGGCGA AGTCGCCGAC CAGGTCCAGC 3000

CGCGCGCCCG CGCCGGGGAG CAGACGGCG AAGCTCTCCG CCAGTTCCCC GACGTCGGGG 3060

ACCTCGCCTT CCCAGGACGC GCGGTGCACG TCCCGGAACG GCTGGGCCCA CTCGGCGGGT 3120

GGCGCGCCCG CGGCCCGCAT CCATTCCGGT GTGCGTCCGG TGGCGCGGGT GAACGCGGGG 3180

TCGTCGAGCA CCTGCCGGGC GGTGGCGTGG TCGGCCACCA CCCACGTCTC GGTGCGGCTG 3240
 CGCCGCACAC CGGACTCGCG CATCGAGCGG TACCGGCGCT GCGGGTCGTC GTCGTGTCCG 3300
 CACAGCAGCA TCGGGTAAGG GTCGCCGTTG CTGCCGTAAC CCCAGTGCAG GCCGCGGATC 3360
 ATCTGGAGCT GCCTGCCCAG CCCGGCGCGA TCGGTCGTGG TCATGAATTC CCTCCGCCCA 3420
 GCCAGGCGTC GATGTGCCG 3439

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 333 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Thr Thr Asp Ala Ala Thr His Val Arg Leu Gly Arg Ser Ala Leu
 1 5 10 15
 Leu Thr Ser Arg Leu Trp Leu Gly Thr Val Asn Phe Ser Gly Arg Val
 20 25 30
 Glu Asp Asp Asp Ala Leu Arg Leu Met Asp His Ala Arg Asp Arg Gly
 35 40 45
 Ile Asn Cys Leu Asp Thr Ala Asp Met Tyr Gly Trp Arg Leu Tyr Lys
 50 55 60
 Gly His Thr Glu Glu Leu Val Gly Arg Trp Leu Ala Gln Gly Gly Gly
 65 70 75 80
 Arg Arg Glu Asp Thr Val Leu Ala Thr Lys Val Gly Gly Glu Met Ser
 85 90 95
 Glu Arg Val Asn Asp Ser Gly Leu Ser Ala Arg His Ile Ile Ala Ser
 100 105 110

Cys Glu Gly Ser Leu Arg Arg Leu Gly Val Asp His Ile Asp Val Tyr
115 120 125

Gln Met His His Ile Asp Arg Ser Ala Pro Trp Asp Glu Val Trp Gln
130 135 140

Ala Met Asp Ser Leu Val Ala Ser Gly Lys Val Ser Tyr Val Gly Ser
145 150 155 160

Ser Asn Phe Ala Gly Trp His Ile Ala Ala Ala Gln Glu Asn Ala Ala
165 170 175

Arg Arg His Ser Leu Gly Met Val Ser His Gln Cys Leu Tyr Asn Leu
180 185 190

Ala Val Arg His Ala Glu Leu Glu Val Leu Pro Ala Ala Gln Ala Tyr
195 200 205

Gly Leu Gly Val Phe Ala Trp Ser Pro Leu His Gly Gly Leu Leu Ser
210 215 220

Gly Ala Leu Glu Lys Leu Ala Ala Gly Thr Ala Val Lys Ser Ala Gln
225 230 235 240

Gly Arg Ala Gln Val Leu Leu Pro Ser Leu Arg Pro Ala Ile Glu Ala
245 250 255

Tyr Glu Lys Phe Cys Arg Asn Leu Gly Glu Asp Pro Ala Glu Val Gly
260 265 270

Leu Ala Trp Val Leu Ser Arg Pro Gly Ile Ala Gly Ala Val Ile Gly
275 280 285

Pro Arg Thr Pro Glu Gln Leu Asp Ser Ala Leu Lys Ala Ser Ala Met
290 295 300

Thr Leu Asp Glu Gln Ala Leu Ser Glu Leu Asp Glu Ile Phe Pro Ala
305 310 315 320

Val Ala Ser Gly Gly Ala Ala Pro Glu Ala Trp Leu Gln
325 330

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 361 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```

Met Thr Thr Thr Asp Arg Ala Gly Leu Gly Arg Gln Leu Gln Met Ile
 1             5             10             15

Arg Gly Leu His Trp Gly Tyr Gly Ser Asn Gly Asp Pro Tyr Pro Met
      20             25             30

Leu Leu Cys Gly His Asp Asp Asp Pro Gln Arg Arg Tyr Arg Ser Met
      35             40             45

Arg Glu Ser Gly Val Arg Arg Ser Arg Thr Glu Thr Trp Val Val Ala
      50             55             60

Asp His Ala Thr Ala Arg Gln Val Leu Asp Asp Pro Ala Phe Thr Arg
      65             70             75             80

Ala Thr Gly Arg Thr Pro Glu Trp Met Arg Ala Ala Gly Ala Pro Pro
      85             90             95

Ala Glu Trp Ala Gln Pro Phe Arg Asp Val His Ala Ala Ser Trp Glu
      100            105            110

Gly Glu Val Pro Asp Val Gly Glu Leu Ala Glu Ser Phe Ala Gly Leu
      115            120            125

Leu Pro Gly Ala Gly Ala Arg Leu Asp Leu Val Gly Asp Phe Ala Trp
      130            135            140

Gln Val Pro Val Gln Gly Met Thr Ala Val Leu Gly Ala Ala Gly Val
      145            150            155            160

Leu Arg Gly Ala Ala Trp Asp Ala Arg Val Ser Leu Asp Ala Gln Leu
      165            170            175

```

Ser Pro Gln Gln Leu Ala Val Thr Glu Ala Ala Val Ala Ala Leu Pro
 180 185 190

Ala Asp Pro Ala Leu Arg Ala Leu Phe Ala Gly Ala Glu Met Thr Ala
 195 200 205

Asn Thr Val Val Asp Ala Val Leu Ala Val Ser Ala Glu Pro Gly Leu
 210 215 220

Ala Glu Arg Ile Ala Asp Asp Pro Ala Ala Ala Gln Arg Thr Val Ala
 225 230 235 240

Glu Val Leu Arg Leu His Pro Ala Leu His Leu Glu Arg Arg Thr Ala
 245 250 255

Thr Ala Glu Val Arg Leu Gly Glu His Val Ile Gly Glu Gly Glu Glu
 260 265 270

Val Val Val Val Val Ala Ala Ala Asn Arg Asp Pro Glu Val Phe Ala
 275 280 285

Glu Pro Asp Arg Leu Asp Val Asp Arg Pro Asp Ala Asp Arg Ala Leu
 290 295 300

Ser Ala His Arg Gly His Pro Gly Arg Leu Glu Glu Leu Val Thr Ala
 305 310 315 320

Leu Ala Thr Ala Ala Leu Arg Ala Ala Ala Lys Ala Leu Pro Gly Leu
 325 330 335

Thr Pro Ser Gly Pro Val Val Arg Arg Arg Arg Ser Pro Val Leu Arg
 340 345 350

Gly Thr Asn Arg Cys Pro Val Glu Leu
 355 360

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1266 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Saccharopolyspora erythraea*

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: complement (4..1266)

(D) OTHER INFORMATION: /function= "involved in the
biosynthesis of desosamine"
/gene= "eryCIII"
/note= "SEQ ID No 1 FROM 1046 TO 2308"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

TCATCGTGGT TCTCTCCTTC CTGCGGCCAG TTCCTCGCAG ATGCCGACGA CCTCGGCCGG	60
TGACGGCTCC GCGAGCATGT CGTCGCGCAT CCGCGCCGCG CCGGCGCGGT GGGCCGGGTC	120
GTCGAGGACC CGCTTCACCG ACTCCCGGAG CTGGTCGGGG GTCAGCTCGG GCACGGGCAG	180
CGCGATCCCC GCCCCGAATT CCTGCGTGCG CTGCGCGCGC ACGCCGGTGT CCCAGCCGTC	240
GGGCAGGATC ACCTGCGGCA CGCCGTGGAT CGCCGCGGTG TGCCAGCTCC CGGGTCCGCC	300
GTGGTGCACC GTCGCCGCGC AGGTCGGCAG CAGCGCGTGC ATCGGGACGA AGCCGACCGT	360
GCGGACGTTG TCCGGGATGT TCGCGACGCC TTCTAGCTGC TCGCGCTCGA AGGTCGCGAT	420
GATCTCGGCG TCGACGTCGC CGACGGCACC CAGCAGCTCC TCGATGGAGA CCTGCCCGAT	480
GCTGTTCTCG CGGCTGGAGA TCCCGAGCGT GAGGCACACG CGGCGGCGCT CGGGCTCGTC	540
GTGCAGCCAT TCCGGCACCA CGGACGGCCC GTTGTAAGTCG ACGTAGCGCA TCCCGACGGT	600
CTTCAGGCCG GTGTCGAGCC TGATCGCGGC CGGGGCGGGG TCGATCGTCC ACTGCCCCGAC	660
GACCACCTCC TCGTCGAAGG CCGGGCCGCC GTACTTCTCC AGCGTCCAGG TGAGCCACTC	720
GGCGAGCGGG TCCTCCCGGT GCTCCTCCGG CTGGTCGGGC AGCAGGCCGA GGAAGTTCTG	780

```

CCGCGCCCGG GTGGTGATGT CGGGTCCCCA CAGCAGCCGC GCGTGCGGCG TTCCGGTCAC      840
CGCCGCCGCG ATGGGCGCGG CGAAGGTGAG CGGCTCCCAG ATGACCAGGT CGGGCCGCCA      900
CTTCCGGCAG AACGAGACCA TGCCTTCGAT GAGCGTGTCC GGGCTCATCA GGGCGTAGAA      960
GGTCGGGGTG AGCACGGTCT GCATGCCCAG CAGGTGCTCC CAGGTCAAGG TGGCGGGGTC     1020
CCGCTCGCTG AAGTCCAGGC TCCGGACGTA GTCGATGATG TCGTGGCCCG CGTGGGTGTC     1080
GAAGTCCACG AGGTGACGCT CGGTGCCGAC CGGGACGGCG GTCAGCCCGG CCGCGGTGAT     1140
GTCCTCGGTG AGCGCCGGGG ACGCGACCAC GCGGACCTCG TGCCCCGCCG CGCGGAACGC     1200
CCATGCGAGG GGGACGAGGC CGAAGAGGTG GCTCTTGCTG GCCATGGAGG AGAAGACGAC     1260
GCGCAT                                                                    1266

```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 421 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```

Met Arg Val Val Phe Ser Ser Met Ala Ser Lys Ser His Leu Phe Gly
 1             5             10             15

Leu Val Pro Leu Ala Trp Ala Phe Arg Ala Ala Gly His Glu Val Arg
      20             25             30

Val Val Ala Ser Pro Ala Leu Thr Glu Asp Ile Thr Ala Ala Gly Leu
      35             40             45

Thr Ala Val Pro Val Gly Thr Asp Val Asp Leu Val Asp Phe Met Thr
      50             55             60

His Ala Gly His Asp Ile Ile Asp Tyr Val Arg Ser Leu Asp Phe Ser

```

65	70	75	80
Glu Arg Asp Pro Ala Thr Leu Thr Trp Glu His Leu Leu Gly Met Gln			
85	90	95	
Thr Val Leu Thr Pro Thr Phe Tyr Ala Leu Met Ser Pro Asp Thr Leu			
100	105	110	
Ile Glu Gly Met Val Ser Phe Cys Arg Lys Trp Arg Pro Asp Leu Val			
115	120	125	
Ile Trp Glu Pro Leu Thr Phe Ala Ala Pro Ile Ala Ala Ala Val Thr			
130	135	140	
Gly Thr Pro His Ala Arg Leu Leu Trp Gly Pro Asp Ile Thr Thr Arg			
145	150	155	160
Ala Arg Gln Asn Phe Leu Gly Leu Leu Pro Asp Gln Pro Glu Glu His			
165	170	175	
Arg Glu Asp Pro Leu Ala Glu Trp Leu Thr Trp Thr Leu Glu Lys Tyr			
180	185	190	
Gly Gly Pro Ala Phe Asp Glu Glu Val Val Val Gly Gln Trp Thr Ile			
195	200	205	
Asp Pro Ala Pro Ala Ala Ile Arg Leu Asp Thr Gly Leu Lys Thr Val			
210	215	220	
Gly Met Arg Tyr Val Asp Tyr Asn Gly Pro Ser Val Val Pro Glu Trp			
225	230	235	240
Leu His Asp Glu Pro Glu Arg Arg Arg Val Cys Leu Thr Leu Gly Ile			
245	250	255	
Ser Ser Arg Glu Asn Ser Ile Gly Gln Val Ser Ile Glu Glu Leu Leu			
260	265	270	
Gly Ala Val Gly Asp Val Asp Ala Glu Ile Ile Ala Thr Phe Asp Ala			
275	280	285	
Gln Gln Leu Glu Gly Val Ala Asn Ile Pro Asp Asn Val Arg Thr Val			
290	295	300	

Gly Phe Val Pro Met His Ala Leu Leu Pro Thr Cys Ala Ala Thr Val
305 310 315 320

His His Gly Gly Pro Gly Ser Trp His Thr Ala Ala Ile His Gly Val
325 330 335

Pro Gln Val Ile Leu Pro Asp Gly Trp Asp Thr Gly Val Arg Ala Gln
340 345 350

Arg Thr Gln Glu Phe Gly Ala Gly Ile Ala Leu Pro Val Pro Glu Leu
355 360 365

Thr Pro Asp Gln Leu Arg Glu Ser Val Lys Arg Val Leu Asp Asp Pro
370 375 380

Ala His Arg Ala Gly Ala Ala Arg Met Arg Asp Asp Met Leu Ala Glu
385 390 395 400

Pro Ser Pro Ala Glu Val Val Gly Ile Cys Glu Glu Leu Ala Ala Gly
405 410 415

Arg Arg Glu Pro Arg
420

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8160 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Saccharopolyspora erythraea

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 242..1207
- (D) OTHER INFORMATION: /function= "involved in the biosynthesis of mycarose"

```
/gene= "eryBIV"  
/transl_except= (pos: 242 .. 244, aa: Met)
```

(ix) FEATURE:

```
(A) NAME/KEY: CDS  
(B) LOCATION:1210..2454  
(D) OTHER INFORMATION:/function= "involved in the  
      biosynthesis of mycarose"  
      /gene= "eryBV"  
      /transl_except= (pos: 1210 .. 1212, aa: Met)
```

(ix) FEATURE:

```
(A) NAME/KEY: CDS  
(B) LOCATION:2510..3220  
(D) OTHER INFORMATION:/function= "involved in the  
      biosynthesis of desosamine"  
      /gene= "eryCVI"
```

(ix) FEATURE:

```
(A) NAME/KEY: CDS  
(B) LOCATION:3308..4837  
(D) OTHER INFORMATION:/function= "involved in the  
      biosynthesis of mycarose"  
      /gene= "eryBVI"  
      /transl_except= (pos: 3308 .. 3310, aa: Met)
```

(ix) FEATURE:

```
(A) NAME/KEY: CDS  
(B) LOCATION:6080..7546  
(D) OTHER INFORMATION:/function= "involved in the  
      biosynthesis of desosamine"  
      /gene= "eryCV"
```

(ix) FEATURE:

```
(A) NAME/KEY: CDS  
(B) LOCATION:7578..8156  
(D) OTHER INFORMATION:/function= "involved in the  
      biosynthesis of mycarose"  
      /gene= "eryBVII"  
      /transl_except= (pos: 7578 .. 7580, aa: Met)
```

(ix) FEATURE:

```
(A) NAME/KEY: mat_peptide
```

(B) LOCATION:242

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

TTTGACAGGT CCGCCACGCG TCCCCCTACT CGACGACCAC GCAATGGGCG AACAATATAG	60
GAAGGATCAA GAGGTTGACA TCGCCTCGTC GAGCCAACGA ACCTGTGAAC ATCTGCATGT	120
TGACAAGATC AACGGCGGCT ACCTACTGTG GTGGCCAGT GACGGGTTGC CGCACATCGC	180
GCTGGGAGAGA TTCTTTGAAT TTCGCCCCTA GCACCGACCT GGAAAGCGAG CAAATGCTCC	240
G GTG AAT GGG ATC AGT GAT TCC CCG CGT CAA TTG ATC ACC CTT CTG	286
Met Asn Gly Ile Ser Asp Ser Pro Arg Gln Leu Ile Thr Leu Leu	
1 5 10 15	
GGC GCT TCC GGC TTC GTC GGG AGC GCG GTT CTG CGC GAG CTG CGC GAC	334
Gly Ala Ser Gly Phe Val Gly Ser Ala Val Leu Arg Glu Leu Arg Asp	
20 25 30	
CAC CCG GTC CGG CTG CGC GCG GTG TCC CGC GGC GGA GCG CCC GCG GTT	382
His Pro Val Arg Leu Arg Ala Val Ser Arg Gly Gly Ala Pro Ala Val	
35 40 45	
CCG CCC GGC GCC GCG GAG GTC GAG GAC CTG CGC GCC GAC CTG CTG GAA	430
Pro Pro Gly Ala Ala Glu Val Glu Asp Leu Arg Ala Asp Leu Leu Glu	
50 55 60	
CCG GGC CGG GCC GCC GCC GCG ATC GAG GAC GCC GAC GTG ATC GTG CAC	478
Pro Gly Arg Ala Ala Ala Ala Ile Glu Asp Ala Asp Val Ile Val His	
65 70 75	
CTG GTG GCG CAC GCA GCG GGC GGT TCC ACC TGG CGC AGC GCC ACC TCC	526
Leu Val Ala His Ala Ala Gly Gly Ser Thr Trp Arg Ser Ala Thr Ser	
80 85 90 95	
GAC CCG GAA GCC GAG CGG GTC AAC GTC GGC CTG ATG CAC GAC CTC GTC	574
Asp Pro Glu Ala Glu Arg Val Asn Val Gly Leu Met His Asp Leu Val	
100 105 110	
GGC GCG CTG CAC GAT CGC CGC AGG TCG ACG CCG CCC GTG TTG CTC TAC	622
Gly Ala Leu His Asp Arg Arg Arg Ser Thr Pro Pro Val Leu Leu Tyr	

115	120	125	
GCG AGC ACC GCA CAG GCC GCG AAC CCG TCG GCG GCC AGC AGG TAC GCG			670
Ala Ser Thr Ala Gln Ala Ala Asn Pro Ser Ala Ala Ser Arg Tyr Ala			
130	135	140	
CAG CAG AAG ACC GAG GCC GAG CGC ATC CTG CGC AAA GCC ACC GAC GAG			718
Gln Gln Lys Thr Glu Ala Glu Arg Ile Leu Arg Lys Ala Thr Asp Glu			
145	150	155	
GGC CGG GTG CGC GGC GTG ATC CTG CGG CTG CCC GCC GTC TAC GGC CAG			766
Gly Arg Val Arg Gly Val Ile Leu Arg Leu Pro Ala Val Tyr Gly Gln			
160	165	170	175
AGC GGC CCG TCC GGC CCC ATG GGG CGG GGC GTG GTC GCA GCG ATG ATC			814
Ser Gly Pro Ser Gly Pro Met Gly Arg Gly Val Val Ala Ala Met Ile			
180	185	190	
CGG CGT GCC CTC GCC GGC GAG CCG CTC ACC ATG TGG CAC GAC GGC GGC			862
Arg Arg Ala Leu Ala Gly Glu Pro Leu Thr Met Trp His Asp Gly Gly			
195	200	205	
GTG CGC CGC GAC CTG CTG CAC GTC GAG GAC GTG GCC ACC GCG TTC GCC			910
Val Arg Arg Asp Leu Leu His Val Glu Asp Val Ala Thr Ala Phe Ala			
210	215	220	
GCC GCG CTG GAG CAC CAC GAC GCG CTG GCC GGC GGC ACG TGG GCG CTG			958
Ala Ala Leu Glu His His Asp Ala Leu Ala Gly Gly Thr Trp Ala Leu			
225	230	235	
GGC GCC GAC CGA TCC GAG CCG CTC GGC GAC ATC TTC CGG GCC GTC TCC			1006
Gly Ala Asp Arg Ser Glu Pro Leu Gly Asp Ile Phe Arg Ala Val Ser			
240	245	250	255
GGC AGC GTC GCC CGG CAG ACC GGC AGC CCC GCC GTC GAC GTG GTC ACC			1054
Gly Ser Val Ala Arg Gln Thr Gly Ser Pro Ala Val Asp Val Val Thr			
260	265	270	
GTG CCC GCG CCC GAG CAC GCC GAG GCC AAC GAC TTC CGC AGC GAC GAC			1102
Val Pro Ala Pro Glu His Ala Glu Ala Asn Asp Phe Arg Ser Asp Asp			
275	280	285	
ATC GAC TCC ACC GAG TTC CGC AGC CGG ACC GGC TGG CGC CCC CGG GTT			1150

Ile Asp Ser Thr Glu Phe Arg Ser Arg Thr Gly Trp Arg Pro Arg Val	
290 295 300	
TCC CTC ACC GAC GGC ATC GAC CGG ACG GTG GCC GCC CTG ACC CCC ACC	1198
Ser Leu Thr Asp Gly Ile Asp Arg Thr Val Ala Ala Leu Thr Pro Thr	
305 310 315	
GAG GAG CAC TA GTG CGG GTA CTG CTG ACG TCC TTC GCG CAC CGC ACG	1245
Glu Glu His Met Arg Val Leu Leu Thr Ser Phe Ala His Arg Thr	
320 1 5 10	
CAC TTC CAG GGA CTG GTC CCG CTG GCG TGG GCG CTG CGC ACC GCG GGT	1293
His Phe Gln Gly Leu Val Pro Leu Ala Trp Ala Leu Arg Thr Ala Gly	
15 20 25	
CAC GAC GTG CGC GTG GCC GCC CAG CCC GCG CTC ACC GAC GCG GTC ATC	1341
His Asp Val Arg Val Ala Ala Gln Pro Ala Leu Thr Asp Ala Val Ile	
30 35 40	
GGC GCC GGT CTC ACC GCG GTA CCC GTC GGC TCC GAC CAC CGG CTG TTC	1389
Gly Ala Gly Leu Thr Ala Val Pro Val Gly Ser Asp His Arg Leu Phe	
45 50 55 60	
GAC ATC GTC CCG GAA GTC GCC GCT CAG GTG CAC CGC TAC TCC TTC TAC	1437
Asp Ile Val Pro Glu Val Ala Ala Gln Val His Arg Tyr Ser Phe Tyr	
65 70 75	
CTG GAC TTC TAC CAC CGC GAG CAG GAG CTG CAC TCG TGG GAG TTC CTG	1485
Leu Asp Phe Tyr His Arg Glu Gln Glu Leu His Ser Trp Glu Phe Leu	
80 85 90	
CTC GGC ATG CAG GAG GCC ACC TCG CGG TGG GTA TAC CCG GTG GTC AAC	1533
Leu Gly Met Gln Glu Ala Thr Ser Arg Trp Val Tyr Pro Val Val Asn	
95 100 105	
AAC GAC TCC TTC GTC GCC GAG CTG GTC GAC TTC GCC CGG GAC TGG CGT	1581
Asn Asp Ser Phe Val Ala Glu Leu Val Asp Phe Ala Arg Asp Trp Arg	
110 115 120	
CCT GAC CTG GTG CTC TGG GAG CCG TTC ACC TTC GCC GGC GCC GTC GCG	1629
Pro Asp Leu Val Leu Trp Glu Pro Phe Thr Phe Ala Gly Ala Val Ala	
125 130 135 140	

GCC CGG GCC TGC GGA GCC GCG CAC GCC CGG CTG CTG TGG GGC AGC GAC	1677
Ala Arg Ala Cys Gly Ala Ala His Ala Arg Leu Leu Trp Gly Ser Asp	
145 150 155	
CTC ACC GGC TAC TTC CGC GGC CGG TTC CAG GCG CAA CGC CTG CGA CGG	1725
Leu Thr Gly Tyr Phe Arg Gly Arg Phe Gln Ala Gln Arg Leu Arg Arg	
160 165 170	
CCG CCG GAG GAC CGG CCG GAC CCG CTG GGC ACG TGG CTG ACC GAG GTC	1773
Pro Pro Glu Asp Arg Pro Asp Pro Leu Gly Thr Trp Leu Thr Glu Val	
175 180 185	
GCG GGG CGC TTC GGC GTC GAA TTC GGC GAG GAC CTC GCG GTC GGG CAG	1821
Ala Gly Arg Phe Gly Val Glu Phe Gly Glu Asp Leu Ala Val Gly Gln	
190 195 200	
TGG TCG GTC GAC CAG TTG CCG CCG AGT TTC CGG CTG GAC ACC GGA ATG	1869
Trp Ser Val Asp Gln Leu Pro Pro Ser Phe Arg Leu Asp Thr Gly Met	
205 210 215 220	
GAA ACC GTT GTC GCG CGG ACC CTG CCC TAC AAC GGC GCG TCG GTG GTT	1917
Glu Thr Val Val Ala Arg Thr Leu Pro Tyr Asn Gly Ala Ser Val Val	
225 230 235	
CCG GAC TGG CTC AAG AAG GGC AGT GCG ACT CGA CGC ATC TGC ATT ACC	1965
Pro Asp Trp Leu Lys Lys Gly Ser Ala Thr Arg Arg Ile Cys Ile Thr	
240 245 250	
GGA GGG TTC TCC GGA CTC GGG CTC GCC GCC GAT GCC GAT CAG TTC GCG	2013
Gly Gly Phe Ser Gly Leu Gly Leu Ala Ala Asp Ala Asp Gln Phe Ala	
255 260 265	
CGG ACG CTC GCG CAG CTC GCG CGA TTC GAT GGC GAA ATC GTG GTT ACG	2061
Arg Thr Leu Ala Gln Leu Ala Arg Phe Asp Gly Glu Ile Val Val Thr	
270 275 280	
GGT TCC GGT CCG GAT ACC TCC GCG GTA CCG GAC AAC ATT CGT TTG GTG	2109
Gly Ser Gly Pro Asp Thr Ser Ala Val Pro Asp Asn Ile Arg Leu Val	
285 290 295 300	
GAT TTC GTT CCG ATG GGC GTT CTG CTC CAG AAC TGC GCG GCG ATC ATC	2157
Asp Phe Val Pro Met Gly Val Leu Leu Gln Asn Cys Ala Ala Ile Ile	
305 310 315	

CAC CAC GGC GGG GCC GGA ACC TGG GCC ACG GCA CTG CAC CAC GGA ATT	2205
His His Gly Gly Ala Gly Thr Trp Ala Thr Ala Leu His His Gly Ile	
320 325 330	
CCG CAA ATA TCA GTT GCA CAT GAA TGG GAT TGC ATG CTA CGC GGC CAG	2253
Pro Gln Ile Ser Val Ala His Glu Trp Asp Cys Met Leu Arg Gly Gln	
335 340 345	
CAG ACC GCG GAA CTG GGC GCG GGA ATC TAC CTC CGG CCG GAC GAG GTC	2301
Gln Thr Ala Glu Leu Gly Ala Gly Ile Tyr Leu Arg Pro Asp Glu Val	
350 355 360	
GAT GCC GAC TCA TTG GCG AGC GCC CTC ACC CAG GTG GTC GAG GAC CCC	2349
Asp Ala Asp Ser Leu Ala Ser Ala Leu Thr Gln Val Val Glu Asp Pro	
365 370 375 380	
ACC TAC ACC GAG AAC GCG GTG AAG CTT CGC GAG GAG GCG CTG TCC GAC	2397
Thr Tyr Thr Glu Asn Ala Val Lys Leu Arg Glu Glu Ala Leu Ser Asp	
385 390 395	
CCG ACG CCG CAG GAG ATC GTC CCG CGA CTG GAG GAA CTC ACG CGC CGC	2445
Pro Thr Pro Gln Glu Ile Val Pro Arg Leu Glu Glu Leu Thr Arg Arg	
400 405 410	
CAC GCC GGC TAGCGGTTTC CGACCGACAA GTCCGTCCGA CAGCACACCT	2494
His Ala Gly	
415	
CCGGAGGGAG CAGGG ATG TAC GAG GGC GGG TTC GCC GAG CTT TAC GAC CGG	2545
Met Tyr Glu Gly Gly Phe Ala Glu Leu Tyr Asp Arg	
1 5 10	
TTC TAC CGC GGC CGG GGC AAG GAC TAC GCG GCC GAG GCC GCG CAG GTC	2593
Phe Tyr Arg Gly Arg Gly Lys Asp Tyr Ala Ala Glu Ala Ala Gln Val	
15 20 25	
GCG CGG CTG GTC AGA GAC CGC CTG CCC TCG GCT TCC TCG CTG CTC GAC	2641
Ala Arg Leu Val Arg Asp Arg Leu Pro Ser Ala Ser Ser Leu Leu Asp	
30 35 40	
GTG GCC TGC GGG ACC GGC ACC CAC CTG CGC CGG TTC GCC GAC CTC TTC	2689
Val Ala Cys Gly Thr Gly Thr His Leu Arg Arg Phe Ala Asp Leu Phe	

45	50	55	60	
GAC GAC GTG ACC GGG CTG GAG CTG TCG GCG GCG ATG ATC GAG GTC GCC				2737
Asp Asp Val Thr Gly Leu Glu Leu Ser Ala Ala Met Ile Glu Val Ala				
65	70	75		
CGG CCG CAG CTC GGC GGC ATC CCG GTG CTG CAG GGC GAC ATG CGC GAC				2785
Arg Pro Gln Leu Gly Gly Ile Pro Val Leu Gln Gly Asp Met Arg Asp				
80	85	90		
TTC GCG CTG GAT CGC GAG TTC GAC GCC GTC ACC TGC ATG TTC AGC TCC				2833
Phe Ala Leu Asp Arg Glu Phe Asp Ala Val Thr Cys Met Phe Ser Ser				
95	100	105		
ATC GGG CAC ATG CGC GAC GGC GCC GAG CTG GAC CAG GCG CTG GCG TCC				2881
Ile Gly His Met Arg Asp Gly Ala Glu Leu Asp Gln Ala Leu Ala Ser				
110	115	120		
TTC GCC CGC CAC CTC GCC CCC GGC GGC GTC GTG GTG GTC GAA CCG TGG				2929
Phe Ala Arg His Leu Ala Pro Gly Gly Val Val Val Val Glu Pro Trp				
125	130	135	140	
TGG TTC CCG GAG GAC TTC CTC GAC GGC TAC GTG GCC GGT GAC GTG GTG				2977
Trp Phe Pro Glu Asp Phe Leu Asp Gly Tyr Val Ala Gly Asp Val Val				
145	150	155		
CGC GAC GGC GAC CTG ACG ATC TCG CGC GTC TCG CAC TCC GTG CGC GCC				3025
Arg Asp Gly Asp Leu Thr Ile Ser Arg Val Ser His Ser Val Arg Ala				
160	165	170		
GGC GGC GCG ACC CGG ATG GAG ATC CAC TGG GTC GTG GCC GAC GCG GTG				3073
Gly Gly Ala Thr Arg Met Glu Ile His Trp Val Val Ala Asp Ala Val				
175	180	185		
AAC GGT CCG CGG CAC CAC GTG GAG CAC TAC GAG ATC ACG CTC TTC GAG				3121
Asn Gly Pro Arg His His Val Glu His Tyr Glu Ile Thr Leu Phe Glu				
190	195	200		
CGG CAG CAG TAC GAG AAG GCC TTC ACC GCG GCC GGT TGC GCT GTG CAG				3169
Arg Gln Gln Tyr Glu Lys Ala Phe Thr Ala Ala Gly Cys Ala Val Gln				
205	210	215	220	
TAC CTG GAG GGC GGA CCC TCC GGA CGC GGG TTG TTC GTC GGT GTG CGC				3217

Tyr Leu Glu Gly Gly Pro Ser Gly Arg Gly Leu Phe Val Gly Val Arg
 225 230 235

GGA TGACCCGTGC GTTCGCGTTT TCCGTTCTCG GCACAGGTGA TCCGCTCCAC 3270
 Gly

GGGCCCTTTC CCCGCCGTGA CCGGACCCTT ACAGTGA GTG CGG GTC TTG ATC GAC 3325
 Met Arg Val Leu Ile Asp
 1 5

AAC GCC CGG CGG CAG CAA GCG GAG CCG TCG ACG ACA CCG CAG GGA GAG 3373
 Asn Ala Arg Arg Gln Gln Ala Glu Pro Ser Thr Thr Pro Gln Gly Gly
 10 15 20

TCG ATG GGT GAT CGG ACC GGC GAC CGG ACG ATT CCG GAA TCC TCG CAG 3421
 Ser Met Gly Asp Arg Thr Gly Asp Arg Thr Ile Pro Glu Ser Ser Gln
 25 30 35

ACC GCA ACG CGT TTC CTG CTC GGC GAC GGC GGA ATC CCC ACC GCC ACG 3469
 Thr Ala Thr Arg Phe Leu Leu Gly Asp Gly Gly Ile Pro Thr Ala Thr
 40 45 50

GCG GAA ACC CAC GAC TGG CTG ACC CGC AAC GGC GCC GAG CAG CGG CTC 3517
 Ala Glu Thr His Asp Trp Leu Thr Arg Asn Gly Ala Glu Gln Arg Leu
 55 60 65 70

GAG GTG GCG CGC GTG CCG TTC AGC GCC ATG GAC CGC TGG TCG TTC CAG 3565
 Glu Val Ala Arg Val Pro Phe Ser Ala Met Asp Arg Trp Ser Phe Gln
 75 80 85

CCC GAG GAC GGC AGG CTC GCC CAC GAG TCC GGG CGC TTC TTC TCC ATC 3613
 Pro Glu Asp Gly Arg Leu Ala His Glu Ser Gly Arg Phe Phe Ser Ile
 90 95 100

GAG GGC CTG CAC GTG CGG ACG AAC TTC GGC TGG CGG CGG GAC TGG ATC 3661
 Glu Gly Leu His Val Arg Thr Asn Phe Gly Trp Arg Arg Asp Trp Ile
 105 110 115

CAG CCC ATC ATC GTG CAG CCC GAG ATC GGC TTC CTC GGC CTC ATC GTC 3709
 Gln Pro Ile Ile Val Gln Pro Glu Ile Gly Phe Leu Gly Leu Ile Val
 120 125 130

AAG	GAG	TTC	GAC	GGT	GTG	CTG	CAC	GTG	CTG	GCG	CAG	GCC	AAG	GCC	GAG	3757
Lys	Glu	Phe	Asp	Gly	Val	Leu	His	Val	Leu	Ala	Gln	Ala	Lys	Ala	Glu	
135					140					145					150	
CCG	GGC	AAC	ATC	AAC	GCC	GTC	CAG	CTC	TCC	CCG	ACC	CTG	CAG	GCG	ACC	3805
Pro	Gly	Asn	Ile	Asn	Ala	Val	Gln	Leu	Ser	Pro	Thr	Leu	Gln	Ala	Thr	
				155					160					165		
CGC	AGC	AAC	TAC	ACC	GGC	GTC	CAC	CGC	GGC	TCG	AAG	GTC	CGG	TTC	ATC	3853
Arg	Ser	Asn	Tyr	Thr	Gly	Val	His	Arg	Gly	Ser	Lys	Val	Arg	Phe	Ile	
			170					175					180			
GAG	TAC	TTC	AAC	GGC	ACG	CGC	CCG	AGC	CGG	ATC	CTC	GTC	GAC	GTG	CTC	3901
Glu	Tyr	Phe	Asn	Gly	Thr	Arg	Pro	Ser	Arg	Ile	Leu	Val	Asp	Val	Leu	
		185					190					195				
CAG	TCC	GAG	CAG	GGC	GCG	TGG	TTC	CTG	CGC	AAG	CGC	AAC	CGG	AAC	ATG	3949
Gln	Ser	Glu	Gln	Gly	Ala	Trp	Phe	Leu	Arg	Lys	Arg	Asn	Arg	Asn	Met	
	200					205				210						
GTC	GTC	GAG	GTG	TTC	GAC	GAC	CTG	CCC	GAG	CAC	CCG	AAC	TTC	CGG	TGG	3997
Val	Val	Glu	Val	Phe	Asp	Asp	Leu	Pro	Glu	His	Pro	Asn	Phe	Arg	Trp	
215					220					225					230	
CTG	ACC	GTC	GCG	CAG	CTG	CGG	GCG	ATG	CTG	CAC	CAC	GAC	AAC	GTG	GTG	4045
Leu	Thr	Val	Ala	Gln	Leu	Arg	Ala	Met	Leu	His	His	Asp	Asn	Val	Val	
			235					240					245			
AAC	ATG	GAC	CTG	CGC	ACC	GTG	CTG	GCC	TGC	GTC	CCG	ACC	GCC	GTG	GAG	4093
Asn	Met	Asp	Leu	Arg	Thr	Val	Leu	Ala	Cys	Val	Pro	Thr	Ala	Val	Glu	
		250					255					260				
CGG	GAC	CGG	GCC	GAC	GAC	GTG	CTC	GCG	CGC	CTG	CCC	GAG	GGC	TCG	TTC	4141
Arg	Asp	Arg	Ala	Asp	Asp	Val	Leu	Ala	Arg	Leu	Pro	Glu	Gly	Ser	Phe	
	265					270					275					
CAG	GCC	CGG	CTG	CTG	CAC	TCG	TTC	ATC	GGC	GCG	GGC	ACC	CCG	GCC	AAC	4189
Gln	Ala	Arg	Leu	Leu	His	Ser	Phe	Ile	Gly	Ala	Gly	Thr	Pro	Ala	Asn	
	280					285				290						
AAC	ATG	AAC	AGC	CTG	CTG	AGC	TGG	ATC	TCC	GAC	GTG	CGC	GCC	AGG	CGC	4237
Asn	Met	Asn	Ser	Leu	Leu	Ser	Trp	Ile	Ser	Asp	Val	Arg	Ala	Arg	Arg	
295				300					305						310	

GAG TTC GTG CAG CGC GGC CGC CCG CTG CCC GAC ATC GAG CGC AGC GGG	4285
Glu Phe Val Gln Arg Gly Arg Pro Leu Pro Asp Ile Glu Arg Ser Gly	
315 320 325	
TGG ATC CGC CGC GAC GAC GGC ATC GAG CAC GAG GAG AAG AAG TAC TTC	4333
Trp Ile Arg Arg Asp Asp Gly Ile Glu His Glu Glu Lys Lys Tyr Phe	
330 335 340	
GAC GTC TTC GGC GTC ACG GTG GCG ACC AGC GAC CGC GAG GTC AAC TCG	4381
Asp Val Phe Gly Val Thr Val Ala Thr Ser Asp Arg Glu Val Asn Ser	
345 350 355	
TGG ATG CAG CCG CTG CTC TCG CCC GCC AAC AAC GGC CTG CTC GCC CTG	4429
Trp Met Gln Pro Leu Leu Ser Pro Ala Asn Asn Gly Leu Leu Ala Leu	
360 365 370	
CTG GTC AAG GAC ATC GGC GGC ACG TTG CAC GCG CTC GTG CAG CTG CGC	4477
Leu Val Lys Asp Ile Gly Gly Thr Leu His Ala Leu Val Gln Leu Arg	
375 380 385 390	
ACC GAG GCG GGC GGG ATG GAC GTC GCC GAG CTG GCG CCT ACG GTG CAC	4525
Thr Glu Ala Gly Gly Met Asp Val Ala Glu Leu Ala Pro Thr Val His	
395 400 405	
TGC CAG CCC GAC AAC TAC GCC GAC GCG CCC GAG GAG TTC CGA CCG GCC	4573
Cys Gln Pro Asp Asn Tyr Ala Asp Ala Pro Glu Glu Phe Arg Pro Ala	
410 415 420	
TAT GTG GAC TAC GTG TTG AAC GTG CCG CGC TCG CAG GTC CGC TAC GAC	4621
Tyr Val Asp Tyr Val Leu Asn Val Pro Arg Ser Gln Val Arg Tyr Asp	
425 430 435	
GCA TGG CAC TCC GAG GAG GGC GGC CGG TTC TAC CGC AAC GAG AAC CGG	4669
Ala Trp His Ser Glu Glu Gly Gly Arg Phe Tyr Arg Asn Glu Asn Arg	
440 445 450	
TAC ATG CTG ATC GAG GTG CCC GCC GAC TTC GAC GCC AGT GCC GCT CCC	4717
Tyr Met Leu Ile Glu Val Pro Ala Asp Phe Asp Ala Ser Ala Ala Pro	
455 460 465 470	
GAC CAC CGG TGG ATG ACC TTC GAC CAG ATC ACC TAC CTG CTC GGG CAC	4765
Asp His Arg Trp Met Thr Phe Asp Gln Ile Thr Tyr Leu Leu Gly His	

475	480	485	
AGC CAC TAC GTC AAC ATC CAG CTG CGC AGC ATC ATC GCG TGC GCC TCG			4813
Ser His Tyr Val Asn Ile Gln Leu Arg Ser Ile Ile Ala Cys Ala Ser			
490	495	500	
GCC GTC TAC ACC AGG ACC GCC GGA TGAAACGCGC GCTGACCGAC CTGGCGATCT			4867
Ala Val Tyr Thr Arg Thr Ala Gly			
505	510		
TCGGCGGCCC CGAGGCATTC CTGCACACCC TCTACGTGGG CAGGCCGACC GTCGGGGACC			4927
GGGAGCGGTT CTTGCCCCGC CTGGAGTGGG CGCTGAACAA CAACTGGCTG ACCAACGGCG			4987
GACCACTGGT GCGCGAGTTC GAGGGCCGGG TCGCCGACCT GGCGGGTGTC CGCCACTGCG			5047
TGGCCACCTG CAACGCGACG GTCGCGCTGC AACTGGTGCT GCGCGCGAGC GACGTGTCCG			5107
GCGAGGTCGT CATGCCTTCG ATGACGTTTCG CGGCCACCGC GCACGCGGCG AGCTGGCTGG			5167
GGCTGGAACC GGTGTTCTGC GACGTGGACC CCGAGACCGG CCTGCTCGAC CCCGAGCAGC			5227
TCGCGTCGCT GGTGACACCG CGGACGGGCG CGATCATCGG CGTGACCTG TGGGGCAGGC			5287
CCGCTCCGGT CGAGGCGCTG GAGAAGATCG CCGCCGAGCA CCAGGTCAAA CTCTTCTTCG			5347
ACGCCGCGCA CGCGCTGGGC TGCACCGCCG GCGGGCGGCC GGTGCGCGCC TTCGGCAACG			5407
CCGAGGTGTT CAGCTTCCAC GCCACGAAGG CGGTCACCTC GTTCGAGGGC GCGCCATCG			5467
TCACCGACGA CGGGCTGCTG GCCGACCGCA TCCGCGCCAT GCACAACTTC GGGATCGCAC			5527
CGGACAAGCT GGTGACCGAT GTCGGCACCA ACGGCAAGAT GAGCGAGTGC GCCGCGGCGA			5587
TGGGCTCAC CTCGCTCGAC GCCTTCGCGG AGACCAGGGT GCACAACCGC CTCAACCACG			5647
CGCTTACTC CGACGAGCTC CGCGACGTGC GCGGCATATC CGTGACGCG TTCGATCCTG			5707
GCGAGCAGAA CAACTACCAG TACGTGATCA TCTCGGTGGA CTCCGCGGCC ACCGGCATCG			5767
ACCGCGACCA GTTGCAGGCG ATCCTGCGAG CGGAGAAGGT TGTGGCACAA CCCTACTTCT			5827
CCCCCGGGT CCACCAGATG CAGCCGTACC GGACCGAGCC GCCGCTGCGG CTGGAGAACA			5887

CCGAACAGCT CTCCGACCGG GTGCTCGCGC TGCCACCGG CCCC GCGGTG TCCAGCGAGG	5947
ACATCCGGCG GGTGTGCGAC ATCATCCGGC TCGCCGCCAC CAGCGGCGAG CTGATCAACG	6007
CGCAATGGGA CCAGAGGACG CGCAACGGTT CGTGACGACC TGCGCCACAA GTGCCAGGAG	6067
GTTCGCTCCC CG ATG AAC ACA ACT CGT ACG GCA ACC GCC CAG GAA GCG	6115
Met Asn Thr Thr Arg Thr Ala Thr Ala Gln Glu Ala	
1 5 10	
GGG GTC GCC GAC GCG GCG CGC CCG GAC GTC GAC CGG CGG GCG GTC GTG	6163
Gly Val Ala Asp Ala Ala Arg Pro Asp Val Asp Arg Arg Ala Val Val	
15 20 25	
CGG GCG CTG AGC TCG GAG GTC TCC CGC GTC ACC GGC GCC GGT GAC GGT	6211
Arg Ala Leu Ser Ser Glu Val Ser Arg Val Thr Gly Ala Gly Asp Gly	
30 35 40	
GAC GCC GAC GTG CAG GCC GCC CGG CTC GCC GAC CTC GCC GCG CAC TAC	6259
Asp Ala Asp Val Gln Ala Ala Arg Leu Ala Asp Leu Ala Ala His Tyr	
45 50 55 60	
GGG GCG CAC CCG TTC ACG CCG CTG GAG CAG ACG CGT GCG CGG CTC GGC	6307
Gly Ala His Pro Phe Thr Pro Leu Glu Gln Thr Arg Ala Arg Leu Gly	
65 70 75	
CTG GAC CGC GCG GAG TTC GCC CAC CTG CTC GAC CTG TTC GGC CGC ATC	6355
Leu Asp Arg Ala Glu Phe Ala His Leu Leu Asp Leu Phe Gly Arg Ile	
80 85 90	
CCG GAC CTG GGC ACC GCG GTG GAG CAC GGT CCG GCG GGC AAG TAC TGG	6403
Pro Asp Leu Gly Thr Ala Val Glu His Gly Pro Ala Gly Lys Tyr Trp	
95 100 105	
TCC AAC ACG ATC AAG CCG CTG GAC GCC GCA GGC GCA CTG GAC GCG GCG	6451
Ser Asn Thr Ile Lys Pro Leu Asp Ala Ala Gly Ala Leu Asp Ala Ala	
110 115 120	
GTC TAC CGC AAG CCT GCC TTC CCC TAC AGC GTC GGC CTG TAC CCC GGG	6499
Val Tyr Arg Lys Pro Ala Phe Pro Tyr Ser Val Gly Leu Tyr Pro Gly	
125 130 135 140	

CCG ACG TGC ATG TTC CGC TGC CAC TTC TGC GTG CGG GTG ACC GGT GCC	6547
Pro Thr Cys Met Phe Arg Cys His Phe Cys Val Arg Val Thr Gly Ala	
145 150 155	
CGC TAC GAG GCC GCA TCG GTC CCG GCG GGC AAC GAG ACG CTG GCC GCG	6595
Arg Tyr Glu Ala Ala Ser Val Pro Ala Gly Asn Glu Thr Leu Ala Ala	
160 165 170	
ATC ATC GAC GAG GTG CCC ACG GAC AAC CCG AAG GCG ATG TAC ATG TCG	6643
Ile Ile Asp Glu Val Pro Thr Asp Asn Pro Lys Ala Met Tyr Met Ser	
175 180 185	
GGC GGG CTC GAG CCG CTG ACC AAC CCC GGT CTC GGC GAG CTG GTG TCG	6691
Gly Gly Leu Glu Pro Leu Thr Asn Pro Gly Leu Gly Glu Leu Val Ser	
190 195 200	
CAC GCC GCC GGG CGC GGT TTC GAC CTC ACC GTC TAC ACC AAC GCC TTC	6739
His Ala Ala Gly Arg Gly Phe Asp Leu Thr Val Tyr Thr Asn Ala Phe	
205 210 215 220	
GCC CTC ACC GAG CAG ACG CTG AAC CGC CAG CCC GGC CTG TGG GAG CTG	6787
Ala Leu Thr Glu Gln Thr Leu Asn Arg Gln Pro Gly Leu Trp Glu Leu	
225 230 235	
GGC GCG ATC CGC ACG TCC CTC TAC GGG CTG AAC AAC GAC GAG TAC GAG	6835
Gly Ala Ile Arg Thr Ser Leu Tyr Gly Leu Asn Asn Asp Glu Tyr Glu	
240 245 250	
ACG ACC ACC GGC AAG CGC GGC GCT TTC GAA CGC GTC AAG AAG AAC CTG	6883
Thr Thr Thr Gly Lys Arg Gly Ala Phe Glu Arg Val Lys Lys Asn Leu	
255 260 265	
CAG GGC TTC CTG CGG ATG CGC GCC GAG CGG GAC GCG CCG ATC CGG CTC	6931
Gln Gly Phe Leu Arg Met Arg Ala Glu Arg Asp Ala Pro Ile Arg Leu	
270 275 280	
GGC TTC AAC CAC ATC ATC CTG CCG GGA CGG GCC GAC CGG CTC ACC GAC	6979
Gly Phe Asn His Ile Ile Leu Pro Gly Arg Ala Asp Arg Leu Thr Asp	
285 290 295 300	
CTC GTC GAC TTC ATC GCC GAG CTC AAC GAG TCC AGC CCG CAA CGG CCG	7027
Leu Val Asp Phe Ile Ala Glu Leu Asn Glu Ser Ser Pro Gln Arg Pro	
305 310 315	

CTG GAC TTC GTG ACG GTG CGC GAG GAC TAC AGC GGC CGC GAC GAC GGC	7075
Leu Asp Phe Val Thr Val Arg Glu Asp Tyr Ser Gly Arg Asp Asp Gly	
320 325 330	
CGG CTG TCG GAC TCC GAG CGC AAC GAG CTG CGC GAG GGC CTG GTG CGG	7123
Arg Leu Ser Asp Ser Glu Arg Asn Glu Leu Arg Glu Gly Leu Val Arg	
335 340 345	
TTC GTC GAC TAC GCC GCC GAG CGG ACC CCG GGC ATG CAC ATC GAC CTG	7171
Phe Val Asp Tyr Ala Ala Glu Arg Thr Pro Gly Met His Ile Asp Leu	
350 355 360	
GGC TAC GCC CTG GAG AGC CTG CGG CGG GGT GTG GAC GCC GAG CTG CTG	7219
Gly Tyr Ala Leu Glu Ser Leu Arg Arg Gly Val Asp Ala Glu Leu Leu	
365 370 375 380	
CGC ATC CGG CCG GAG ACG ATG CGT CCC ACC GCG CAC CCC CAG GTC GCG	7267
Arg Ile Arg Pro Glu Thr Met Arg Pro Thr Ala His Pro Gln Val Ala	
385 390 395	
GTG CAG ATC GAC CTG CTC GGC GAC GTC TAC CTC TAC CGC GAG GCG GGC	7315
Val Gln Ile Asp Leu Leu Gly Asp Val Tyr Leu Tyr Arg Glu Ala Gly	
400 405 410	
TTC CCG GAG CTG GAG GGC GCC ACC CGC TAC ATC GCG GGC CGG GTC ACC	7363
Phe Pro Glu Leu Glu Gly Ala Thr Arg Tyr Ile Ala Gly Arg Val Thr	
415 420 425	
CCG TCG ACC AGC CTG CGC GAG GTG GTG GAG AAC TTC GTG CTG GAG AAC	7411
Pro Ser Thr Ser Leu Arg Glu Val Val Glu Asn Phe Val Leu Glu Asn	
430 435 440	
GAG GGC GTG CAG CCC CGC CCC GGC GAC GAG TAC TTC CTC GAC GGC TTC	7459
Glu Gly Val Gln Pro Arg Pro Gly Asp Glu Tyr Phe Leu Asp Gly Phe	
445 450 455 460	
GAC CAG TCG GTG ACC GCA CGG CTC AAC CAG CTC GAA CGA GAC ATC GCC	7507
Asp Gln Ser Val Thr Ala Arg Leu Asn Gln Leu Glu Arg Asp Ile Ala	
465 470 475	
GAC GGG TGG GAG GAC CAC CGC GGC TTC CTG CGC GGA AGG TGAACCGGAG	7556
Asp Gly Trp Glu Asp His Arg Gly Phe Leu Arg Gly Arg	

480	485	
TTGCGAGTAC GTGAGCTGGC G GTG GCG GGC GGT TTC GAG TTC ACC CCC GAC		7607
	Met Ala Gly Gly Phe Glu Phe Thr Pro Asp	
	1 5 10	
CCG AAG CAG GAC CGG CGG GGC CTG TTC GTG TCT CCG CTG CAG GAC GAG		7655
Pro Lys Gln Asp Arg Arg Gly Leu Phe Val Ser Pro Leu Gln Asp Glu		
	15 20 25	
GCG TTC GTG GGC GCG GTG GGC CAT CGG TTC CCC GTC GCC CAG ATG AAC		7703
Ala Phe Val Gly Ala Val Gly His Arg Phe Pro Val Ala Gln Met Asn		
	30 35 40	
CAC ATC GTC TCC GCC CGG GGC GTG CTG CGC GGG CTG CAC TTC ACC ACC		7751
His Ile Val Ser Ala Arg Gly Val Leu Arg Gly Leu His Phe Thr Thr		
	45 50 55	
ACC CCG CCG GGG CAG TGC AAG TAC GTC TAC TGC GCG CGC GGC CGG GCG		7799
Thr Pro Pro Gly Gln Cys Lys Tyr Val Tyr Cys Ala Arg Gly Arg Ala		
	60 65 70	
CTC GAC GTC ATC GTC GAC ATC CGG GTC GGC TCG CCG ACG TTC GGG AAG		7847
Leu Asp Val Ile Val Asp Ile Arg Val Gly Ser Pro Thr Phe Gly Lys		
	75 80 85 90	
TGG GAC GCG GTG GAG ATG GAC ACC GAG CAC TTC CGG GCG GTC TAC TTC		7895
Trp Asp Ala Val Glu Met Asp Thr Glu His Phe Arg Ala Val Tyr Phe		
	95 100 105	
CCC AGG GGC ACC GCG CAC GCC TTC CTC GCG CTT GAG GAC GAC ACC CTG		7943
Pro Arg Gly Thr Ala His Ala Phe Leu Ala Leu Glu Asp Asp Thr Leu		
	110 115 120	
ATG TCG TAC CTG GTC AGC ACG CCG TAC GTG GCC GAG TAC GAG CAG GCG		7991
Met Ser Tyr Leu Val Ser Thr Pro Tyr Val Ala Glu Tyr Glu Gln Ala		
	125 130 135	
ATC GAC CCG TTC GAC CCC GCG CTG GGT CTG CCG TGG CCC GCG GAC CTG		8039
Ile Asp Pro Phe Asp Pro Ala Leu Gly Leu Pro Trp Pro Ala Asp Leu		
	140 145 150	
GAG GTC GTG CTC TCC GAC CGC GAC ACG GTG GCC GTG GAC CTG GAG ACC		8087

Glu Val Val Leu Ser Asp Arg Asp Thr Val Ala Val Asp Leu Glu Thr
 155 160 165 170

GCC AGG CGG CGA GGG ATG CTG CCC GAC TAC GCC GAC TGC CTC GGC GAG 8135
 Ala Arg Arg Arg Gly Met Leu Pro Asp Tyr Ala Asp Cys Leu Gly Glu
 175 180 185

GAG CCC GCC AGC ACC GGC AGG TGAC 8160
 Glu Pro Ala Ser Thr Gly Arg
 190

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 322 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

Met Asn Gly Ile Ser Asp Ser Pro Arg Gln Leu Ile Thr Leu Leu Gly
 1 5 10 15

Ala Ser Gly Phe Val Gly Ser Ala Val Leu Arg Glu Leu Arg Asp His
 20 25 30

Pro Val Arg Leu Arg Ala Val Ser Arg Gly Gly Ala Pro Ala Val Pro
 35 40 45

Pro Gly Ala Ala Glu Val Glu Asp Leu Arg Ala Asp Leu Leu Glu Pro
 50 55 60

Gly Arg Ala Ala Ala Ala Ile Glu Asp Ala Asp Val Ile Val His Leu
 65 70 75 80

Val Ala His Ala Ala Gly Gly Ser Thr Trp Arg Ser Ala Thr Ser Asp
 85 90 95

Pro Glu Ala Glu Arg Val Asn Val Gly Leu Met His Asp Leu Val Gly
 100 105 110

Ala Leu His Asp Arg Arg Arg Ser Thr Pro Pro Val Leu Leu Tyr Ala
115 120 125

Ser Thr Ala Gln Ala Ala Asn Pro Ser Ala Ala Ser Arg Tyr Ala Gln
130 135 140

Gln Lys Thr Glu Ala Glu Arg Ile Leu Arg Lys Ala Thr Asp Glu Gly
145 150 155 160

Arg Val Arg Gly Val Ile Leu Arg Leu Pro Ala Val Tyr Gly Gln Ser
165 170 175

Gly Pro Ser Gly Pro Met Gly Arg Gly Val Val Ala Ala Met Ile Arg
180 185 190

Arg Ala Leu Ala Gly Glu Pro Leu Thr Met Trp His Asp Gly Gly Val
195 200 205

Arg Arg Asp Leu Leu His Val Glu Asp Val Ala Thr Ala Phe Ala Ala
210 215 220

Ala Leu Glu His His Asp Ala Leu Ala Gly Gly Thr Trp Ala Leu Gly
225 230 235 240

Ala Asp Arg Ser Glu Pro Leu Gly Asp Ile Phe Arg Ala Val Ser Gly
245 250 255

Ser Val Ala Arg Gln Thr Gly Ser Pro Ala Val Asp Val Val Thr Val
260 265 270

Pro Ala Pro Glu His Ala Glu Ala Asn Asp Phe Arg Ser Asp Asp Ile
275 280 285

Asp Ser Thr Glu Phe Arg Ser Arg Thr Gly Trp Arg Pro Arg Val Ser
290 295 300

Leu Thr Asp Gly Ile Asp Arg Thr Val Ala Ala Leu Thr Pro Thr Glu
305 310 315 320

Glu His

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 415 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```

Met Arg Val Leu Leu Thr Ser Phe Ala His Arg Thr His Phe Gln Gly
 1             5             10             15

Leu Val Pro Leu Ala Trp Ala Leu Arg Thr Ala Gly His Asp Val Arg
          20             25             30

Val Ala Ala Gln Pro Ala Leu Thr Asp Ala Val Ile Gly Ala Gly Leu
          35             40             45

Thr Ala Val Pro Val Gly Ser Asp His Arg Leu Phe Asp Ile Val Pro
          50             55             60

Glu Val Ala Ala Gln Val His Arg Tyr Ser Phe Tyr Leu Asp Phe Tyr
          65             70             75             80

His Arg Glu Gln Glu Leu His Ser Trp Glu Phe Leu Leu Gly Met Gln
          85             90             95

Glu Ala Thr Ser Arg Trp Val Tyr Pro Val Val Asn Asn Asp Ser Phe
          100            105            110

Val Ala Glu Leu Val Asp Phe Ala Arg Asp Trp Arg Pro Asp Leu Val
          115            120            125

Leu Trp Glu Pro Phe Thr Phe Ala Gly Ala Val Ala Ala Arg Ala Cys
          130            135            140

Gly Ala Ala His Ala Arg Leu Leu Trp Gly Ser Asp Leu Thr Gly Tyr
          145            150            155            160

Phe Arg Gly Arg Phe Gln Ala Gln Arg Leu Arg Arg Pro Pro Glu Asp
          165            170            175

```

Arg Pro Asp Pro Leu Gly Thr Trp Leu Thr Glu Val Ala Gly Arg Phe
 180 185 190

Gly Val Glu Phe Gly Glu Asp Leu Ala Val Gly Gln Trp Ser Val Asp
 195 200 205

Gln Leu Pro Pro Ser Phe Arg Leu Asp Thr Gly Met Glu Thr Val Val
 210 215 220

Ala Arg Thr Leu Pro Tyr Asn Gly Ala Ser Val Val Pro Asp Trp Leu
 225 230 235 240

Lys Lys Gly Ser Ala Thr Arg Arg Ile Cys Ile Thr Gly Gly Phe Ser
 245 250 255

Gly Leu Gly Leu Ala Ala Asp Ala Asp Gln Phe Ala Arg Thr Leu Ala
 260 265 270

Gln Leu Ala Arg Phe Asp Gly Glu Ile Val Val Thr Gly Ser Gly Pro
 275 280 285

Asp Thr Ser Ala Val Pro Asp Asn Ile Arg Leu Val Asp Phe Val Pro
 290 295 300

Met Gly Val Leu Leu Gln Asn Cys Ala Ala Ile Ile His His Gly Gly
 305 310 315 320

Ala Gly Thr Trp Ala Thr Ala Leu His His Gly Ile Pro Gln Ile Ser
 325 330 335

Val Ala His Glu Trp Asp Cys Met Leu Arg Gly Gln Gln Thr Ala Glu
 340 345 350

Leu Gly Ala Gly Ile Tyr Leu Arg Pro Asp Glu Val Asp Ala Asp Ser
 355 360 365

Leu Ala Ser Ala Leu Thr Gln Val Val Glu Asp Pro Thr Tyr Thr Glu
 370 375 380

Asn Ala Val Lys Leu Arg Glu Glu Ala Leu Ser Asp Pro Thr Pro Gln
 385 390 395 400

Glu Ile Val Pro Arg Leu Glu Glu Leu Thr Arg Arg His Ala Gly

405

410

415

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 237 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

```

Met Tyr Glu Gly Gly Phe Ala Glu Leu Tyr Asp Arg Phe Tyr Arg Gly
 1             5             10             15

Arg Gly Lys Asp Tyr Ala Ala Glu Ala Ala Gln Val Ala Arg Leu Val
      20             25             30

Arg Asp Arg Leu Pro Ser Ala Ser Ser Leu Leu Asp Val Ala Cys Gly
      35             40             45

Thr Gly Thr His Leu Arg Arg Phe Ala Asp Leu Phe Asp Asp Val Thr
      50             55             60

Gly Leu Glu Leu Ser Ala Ala Met Ile Glu Val Ala Arg Pro Gln Leu
      65             70             75             80

Gly Gly Ile Pro Val Leu Gln Gly Asp Met Arg Asp Phe Ala Leu Asp
      85             90             95

Arg Glu Phe Asp Ala Val Thr Cys Met Phe Ser Ser Ile Gly His Met
      100             105             110

Arg Asp Gly Ala Glu Leu Asp Gln Ala Leu Ala Ser Phe Ala Arg His
      115             120             125

Leu Ala Pro Gly Gly Val Val Val Val Glu Pro Trp Trp Phe Pro Glu
      130             135             140

Asp Phe Leu Asp Gly Tyr Val Ala Gly Asp Val Val Arg Asp Gly Asp
      145             150             155             160

```



```

Leu Thr Ile Ser Arg Val Ser His Ser Val Arg Ala Gly Gly Ala Thr
      165                      170                      175

Arg Met Glu Ile His Trp Val Val Ala Asp Ala Val Asn Gly Pro Arg
      180                      185                      190

His His Val Glu His Tyr Glu Ile Thr Leu Phe Glu Arg Gln Gln Tyr
      195                      200                      205

Glu Lys Ala Phe Thr Ala Ala Gly Cys Ala Val Gln Tyr Leu Glu Gly
      210                      215                      220

Gly Pro Ser Gly Arg Gly Leu Phe Val Gly Val Arg Gly
      225                      230                      235

```

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 510 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

```

Met Arg Val Leu Ile Asp Asn Ala Arg Arg Gln Gln Ala Glu Pro Ser
  1              5              10              15

Thr Thr Pro Gln Gly Glu Ser Met Gly Asp Arg Thr Gly Asp Arg Thr
      20              25              30

Ile Pro Glu Ser Ser Gln Thr Ala Thr Arg Phe Leu Leu Gly Asp Gly
      35              40              45

Gly Ile Pro Thr Ala Thr Ala Glu Thr His Asp Trp Leu Thr Arg Asn
      50              55              60

Gly Ala Glu Gln Arg Leu Glu Val Ala Arg Val Pro Phe Ser Ala Met
      65              70              75              80

Asp Arg Trp Ser Phe Gln Pro Glu Asp Gly Arg Leu Ala His Glu Ser
      85              90              95

```

Gly Arg Phe Phe Ser Ile Glu Gly Leu His Val Arg Thr Asn Phe Gly
100 105 110

Trp Arg Arg Asp Trp Ile Gln Pro Ile Ile Val Gln Pro Glu Ile Gly
115 120 125

Phe Leu Gly Leu Ile Val Lys Glu Phe Asp Gly Val Leu His Val Leu
130 135 140

Ala Gln Ala Lys Ala Glu Pro Gly Asn Ile Asn Ala Val Gln Leu Ser
145 150 155 160

Pro Thr Leu Gln Ala Thr Arg Ser Asn Tyr Thr Gly Val His Arg Gly
165 170 175

Ser Lys Val Arg Phe Ile Glu Tyr Phe Asn Gly Thr Arg Pro Ser Arg
180 185 190

Ile Leu Val Asp Val Leu Gln Ser Glu Gln Gly Ala Trp Phe Leu Arg
195 200 205

Lys Arg Asn Arg Asn Met Val Val Glu Val Phe Asp Asp Leu Pro Glu
210 215 220

His Pro Asn Phe Arg Trp Leu Thr Val Ala Gln Leu Arg Ala Met Leu
225 230 235 240

His His Asp Asn Val Val Asn Met Asp Leu Arg Thr Val Leu Ala Cys
245 250 255

Val Pro Thr Ala Val Glu Arg Asp Arg Ala Asp Asp Val Leu Ala Arg
260 265 270

Leu Pro Glu Gly Ser Phe Gln Ala Arg Leu Leu His Ser Phe Ile Gly
275 280 285

Ala Gly Thr Pro Ala Asn Asn Met Asn Ser Leu Leu Ser Trp Ile Ser
290 295 300

Asp Val Arg Ala Arg Arg Glu Phe Val Gln Arg Gly Arg Pro Leu Pro
305 310 315 320

```

Asp Ile Glu Arg Ser Gly Trp Ile Arg Arg Asp Asp Gly Ile Glu His
      325                      330                      335

Glu Glu Lys Lys Tyr Phe Asp Val Phe Gly Val Thr Val Ala Thr Ser
      340                      345                      350

Asp Arg Glu Val Asn Ser Trp Met Gln Pro Leu Leu Ser Pro Ala Asn
      355                      360                      365

Asn Gly Leu Leu Ala Leu Leu Val Lys Asp Ile Gly Gly Thr Leu His
      370                      375                      380

Ala Leu Val Gln Leu Arg Thr Glu Ala Gly Gly Met Asp Val Ala Glu
      385                      390                      395                      400

Leu Ala Pro Thr Val His Cys Gln Pro Asp Asn Tyr Ala Asp Ala Pro
      405                      410                      415

Glu Glu Phe Arg Pro Ala Tyr Val Asp Tyr Val Leu Asn Val Pro Arg
      420                      425                      430

Ser Gln Val Arg Tyr Asp Ala Trp His Ser Glu Glu Gly Gly Arg Phe
      435                      440                      445

Tyr Arg Asn Glu Asn Arg Tyr Met Leu Ile Glu Val Pro Ala Asp Phe
      450                      455                      460

Asp Ala Ser Ala Ala Pro Asp His Arg Trp Met Thr Phe Asp Gln Ile
      465                      470                      475                      480

Thr Tyr Leu Leu Gly His Ser His Tyr Val Asn Ile Gln Leu Arg Ser
      485                      490                      495

Ile Ile Ala Cys Ala Ser Ala Val Tyr Thr Arg Thr Ala Gly
      500                      505                      510

```

(2) INFORMATION FOR SEQ ID NO: 11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 489 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

```

Met Asn Thr Thr Arg Thr Ala Thr Ala Gln Glu Ala Gly Val Ala Asp
 1             5             10             15

Ala Ala Arg Pro Asp Val Asp Arg Arg Ala Val Val Arg Ala Leu Ser
      20             25             30

Ser Glu Val Ser Arg Val Thr Gly Ala Gly Asp Gly Asp Ala Asp Val
      35             40             45

Gln Ala Ala Arg Leu Ala Asp Leu Ala Ala His Tyr Gly Ala His Pro
      50             55             60

Phe Thr Pro Leu Glu Gln Thr Arg Ala Arg Leu Gly Leu Asp Arg Ala
      65             70             75             80

Glu Phe Ala His Leu Leu Asp Leu Phe Gly Arg Ile Pro Asp Leu Gly
      85             90             95

Thr Ala Val Glu His Gly Pro Ala Gly Lys Tyr Trp Ser Asn Thr Ile
      100             105             110

Lys Pro Leu Asp Ala Ala Gly Ala Leu Asp Ala Ala Val Tyr Arg Lys
      115             120             125

Pro Ala Phe Pro Tyr Ser Val Gly Leu Tyr Pro Gly Pro Thr Cys Met
      130             135             140

Phe Arg Cys His Phe Cys Val Arg Val Thr Gly Ala Arg Tyr Glu Ala
      145             150             155             160

Ala Ser Val Pro Ala Gly Asn Glu Thr Leu Ala Ala Ile Ile Asp Glu
      165             170             175

Val Pro Thr Asp Asn Pro Lys Ala Met Tyr Met Ser Gly Gly Leu Glu
      180             185             190

Pro Leu Thr Asn Pro Gly Leu Gly Glu Leu Val Ser His Ala Ala Gly
      195             200             205

```

Arg Gly Phe Asp Leu Thr Val Tyr Thr Asn Ala Phe Ala Leu Thr Glu
 210 215 220

Gln Thr Leu Asn Arg Gln Pro Gly Leu Trp Glu Leu Gly Ala Ile Arg
 225 230 235 240

Thr Ser Leu Tyr Gly Leu Asn Asn Asp Glu Tyr Glu Thr Thr Thr Gly
 245 250 255

Lys Arg Gly Ala Phe Glu Arg Val Lys Lys Asn Leu Gln Gly Phe Leu
 260 265 270

Arg Met Arg Ala Glu Arg Asp Ala Pro Ile Arg Leu Gly Phe Asn His
 275 280 285

Ile Ile Leu Pro Gly Arg Ala Asp Arg Leu Thr Asp Leu Val Asp Phe
 290 295 300

Ile Ala Glu Leu Asn Glu Ser Ser Pro Gln Arg Pro Leu Asp Phe Val
 305 310 315 320

Thr Val Arg Glu Asp Tyr Ser Gly Arg Asp Asp Gly Arg Leu Ser Asp
 325 330 335

Ser Glu Arg Asn Glu Leu Arg Glu Gly Leu Val Arg Phe Val Asp Tyr
 340 345 350

Ala Ala Glu Arg Thr Pro Gly Met His Ile Asp Leu Gly Tyr Ala Leu
 355 360 365

Glu Ser Leu Arg Arg Gly Val Asp Ala Glu Leu Leu Arg Ile Arg Pro
 370 375 380

Glu Thr Met Arg Pro Thr Ala His Pro Gln Val Ala Val Gln Ile Asp
 385 390 395 400

Leu Leu Gly Asp Val Tyr Leu Tyr Arg Glu Ala Gly Phe Pro Glu Leu
 405 410 415

Glu Gly Ala Thr Arg Tyr Ile Ala Gly Arg Val Thr Pro Ser Thr Ser
 420 425 430

Leu Arg Glu Val Val Glu Asn Phe Val Leu Glu Asn Glu Gly Val Gln

```

                435                440                445
Pro Arg Pro Gly Asp Glu Tyr Phe Leu Asp Gly Phe Asp Gln Ser Val
  450                455                460

Thr Ala Arg Leu Asn Gln Leu Glu Arg Asp Ile Ala Asp Gly Trp Glu
  465                470                475                480

Asp His Arg Gly Phe Leu Arg Gly Arg
                485

```

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 193 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

```

Met Ala Gly Gly Phe Glu Phe Thr Pro Asp Pro Lys Gln Asp Arg Arg
  1      5      10      15

Gly Leu Phe Val Ser Pro Leu Gln Asp Glu Ala Phe Val Gly Ala Val
      20      25      30

Gly His Arg Phe Pro Val Ala Gln Met Asn His Ile Val Ser Ala Arg
      35      40      45

Gly Val Leu Arg Gly Leu His Phe Thr Thr Thr Pro Pro Gly Gln Cys
      50      55      60

Lys Tyr Val Tyr Cys Ala Arg Gly Arg Ala Leu Asp Val Ile Val Asp
      65      70      75      80

Ile Arg Val Gly Ser Pro Thr Phe Gly Lys Trp Asp Ala Val Glu Met
      85      90      95

Asp Thr Glu His Phe Arg Ala Val Tyr Phe Pro Arg Gly Thr Ala His
      100      105      110

```

Ala Phe Leu Ala Leu Glu Asp Asp Thr Leu Met Ser Tyr Leu Val Ser
115 120 125

Thr Pro Tyr Val Ala Glu Tyr Glu Gln Ala Ile Asp Pro Phe Asp Pro
130 135 140

Ala Leu Gly Leu Pro Trp Pro Ala Asp Leu Glu Val Val Leu Ser Asp
145 150 155 160

Arg Asp Thr Val Ala Val Asp Leu Glu Thr Ala Arg Arg Arg Gly Met
165 170 175

Leu Pro Asp Tyr Ala Asp Cys Leu Gly Glu Glu Pro Ala Ser Thr Gly
180 185 190

Arg

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1206 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Saccharopolyspora erythraea

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION:1..1203
- (D) OTHER INFORMATION:/function= "involved in the
biosynthesis of desosamine"
/gene= "eryCIV"
/note= "SEQ ID No 6 FROM 4837 TO 6039"

(ix) FEATURE:

- (A) NAME/KEY: mat_peptide
- (B) LOCATION:1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

ATG AAA CGC GCG CTG ACC GAC CTG GCG ATC TTC GGC GGC CCC GAG GCA	48
Met Lys Arg Ala Leu Thr Asp Leu Ala Ile Phe Gly Gly Pro Glu Ala	
1 5 10 15	
TTC CTG CAC ACC CTC TAC GTG GGC AGG CCG ACC GTC GGG GAC CGG GAG	96
Phe Leu His Thr Leu Tyr Val Gly Arg Pro Thr Val Gly Asp Arg Glu	
20 25 30	
CGG TTC TTC GCC CGC CTG GAG TGG GCG CTG AAC AAC AAC TGG CTG ACC	144
Arg Phe Phe Ala Arg Leu Glu Trp Ala Leu Asn Asn Asn Trp Leu Thr	
35 40 45	
AAC GGC GGA CCA CTG GTG CGC GAG TTC GAG GGC CGG GTC GCC GAC CTG	192
Asn Gly Gly Pro Leu Val Arg Glu Phe Glu Gly Arg Val Ala Asp Leu	
50 55 60	
GCG GGT GTC CGC CAC TGC GTG GCC ACC TGC AAC GCG ACG GTC GCG CTG	240
Ala Gly Val Arg His Cys Val Ala Thr Cys Asn Ala Thr Val Ala Leu	
65 70 75 80	
CAA CTG GTG CTG CGC GCG AGC GAC GTG TCC GGC GAG GTC GTC ATG CCT	288
Gln Leu Val Leu Arg Ala Ser Asp Val Ser Gly Glu Val Val Met Pro	
85 90 95	
TCG ATG ACG TTC GCG GCC ACC GCG CAC GCG GCG AGC TGG CTG GGG CTG	336
Ser Met Thr Phe Ala Ala Thr Ala His Ala Ala Ser Trp Leu Gly Leu	
100 105 110	
GAA CCG GTG TTC TGC GAC GTG GAC CCC GAG ACC GGC CTG CTC GAC CCC	384
Glu Pro Val Phe Cys Asp Val Asp Pro Glu Thr Gly Leu Leu Asp Pro	
115 120 125	
GAG CAC GTC GCG TCG CTG GTG ACA CCG CGG ACG GGC GCG ATC ATC GGC	432
Glu His Val Ala Ser Leu Val Thr Pro Arg Thr Gly Ala Ile Ile Gly	
130 135 140	
GTG CAC CTG TGG GGC AGG CCC GCT CCG GTC GAG GCG CTG GAG AAG ATC	480
Val His Leu Trp Gly Arg Pro Ala Pro Val Glu Ala Leu Glu Lys Ile	
145 150 155 160	

GCC GCC GAG CAC CAG GTC AAA CTC TTC TTC GAC GCC GCG CAC GCG CTG	528
Ala Ala Glu His Gln Val Lys Leu Phe Phe Asp Ala Ala His Ala Leu	
165 170 175	
GGC TGC ACC GCC GGC GGG CGG CCG GTC GGC GCC TTC GGC AAC GCC GAG	576
Gly Cys Thr Ala Gly Gly Arg Pro Val Gly Ala Phe Gly Asn Ala Glu	
180 185 190	
GTG TTC AGC TTC CAC GCC ACG AAG GCG GTC ACC TCG TTC GAG GGC GGC	624
Val Phe Ser Phe His Ala Thr Lys Ala Val Thr Ser Phe Glu Gly Gly	
195 200 205	
GCC ATC GTC ACC GAC GAC GGG CTG CTG GCC GAC CGC ATC CGC GCC ATG	672
Ala Ile Val Thr Asp Asp Gly Leu Leu Ala Asp Arg Ile Arg Ala Met	
210 215 220	
CAC AAC TTC GGG ATC GCA CCG GAC AAG CTG GTG ACC GAT GTC GGC ACC	720
His Asn Phe Gly Ile Ala Pro Asp Lys Leu Val Thr Asp Val Gly Thr	
225 230 235 240	
AAC GGC AAG ATG AGC GAG TGC GCC GCG GCG ATG GGC CTC ACC TCG CTC	768
Asn Gly Lys Met Ser Glu Cys Ala Ala Ala Met Gly Leu Thr Ser Leu	
245 250 255	
GAC GCC TTC GCC GAG ACC AGG GTG CAC AAC CGC CTC AAC CAC GCG CTC	816
Asp Ala Phe Ala Glu Thr Arg Val His Asn Arg Leu Asn His Ala Leu	
260 265 270	
TAC TCC GAC GAG CTC CGC GAC GTG CGC GGC ATA TCC GTG CAC GCG TTC	864
Tyr Ser Asp Glu Leu Arg Asp Val Arg Gly Ile Ser Val His Ala Phe	
275 280 285	
GAT CCT GGC GAG CAG AAC AAC TAC CAG TAC GTG ATC ATC TCG GTG GAC	912
Asp Pro Gly Glu Gln Asn Asn Tyr Gln Tyr Val Ile Ile Ser Val Asp	
290 295 300	
TCC GCG GCC ACC GGC ATC GAC CGC GAC CAG TTG CAG GCG ATC CTG CGA	960
Ser Ala Ala Thr Gly Ile Asp Arg Asp Gln Leu Gln Ala Ile Leu Arg	
305 310 315 320	
GCG GAG AAG GTT GTG GCA CAA CCC TAC TTC TCC CCC GGC TGC CAC CAG	1008
Ala Glu Lys Val Val Ala Gln Pro Tyr Phe Ser Pro Gly Cys His Gln	
325 330 335	

ATG CAG CCG TAC CGG ACC GAG CCG CCG CTG CGG CTG GAG AAC ACC GAA	1056
Met Gln Pro Tyr Arg Thr Glu Pro Pro Leu Arg Leu Glu Asn Thr Glu	
340 345 350	
CAG CTC TCC GAC CGG GTG CTC GCG CTG CCC ACC GGC CCC GCG GTG TCC	1104
Gln Leu Ser Asp Arg Val Leu Ala Leu Pro Thr Gly Pro Ala Val Ser	
355 360 365	
AGC GAG GAC ATC CGG CGG GTG TGC GAC ATC ATC CGG CTC GCC GCC ACC	1152
Ser Glu Asp Ile Arg Arg Val Cys Asp Ile Ile Arg Leu Ala Ala Thr	
370 375 380	
AGC GGC GAG CTG ATC AAC GCG CAA TGG GAC CAG AGG ACG CGC AAC GGT	1200
Ser Gly Glu Leu Ile Asn Ala Gln Trp Asp Gln Arg Thr Arg Asn Gly	
385 390 395 400	
TCG TGA	1206
Ser	

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 401 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Met Lys Arg Ala Leu Thr Asp Leu Ala Ile Phe Gly Gly Pro Glu Ala	
1 5 10 15	
Phe Leu His Thr Leu Tyr Val Gly Arg Pro Thr Val Gly Asp Arg Glu	
20 25 30	
Arg Phe Phe Ala Arg Leu Glu Trp Ala Leu Asn Asn Asn Trp Leu Thr	
35 40 45	
Asn Gly Gly Pro Leu Val Arg Glu Phe Glu Gly Arg Val Ala Asp Leu	
50 55 60	

Ala Gly Val Arg His Cys Val Ala Thr Cys Asn Ala Thr Val Ala Leu
65 70 75 80

Gln Leu Val Leu Arg Ala Ser Asp Val Ser Gly Glu Val Val Met Pro
85 90 95

Ser Met Thr Phe Ala Ala Thr Ala His Ala Ala Ser Trp Leu Gly Leu
100 105 110

Glu Pro Val Phe Cys Asp Val Asp Pro Glu Thr Gly Leu Leu Asp Pro
115 120 125

Glu His Val Ala Ser Leu Val Thr Pro Arg Thr Gly Ala Ile Ile Gly
130 135 140

Val His Leu Trp Gly Arg Pro Ala Pro Val Glu Ala Leu Glu Lys Ile
145 150 155 160

Ala Ala Glu His Gln Val Lys Leu Phe Phe Asp Ala Ala His Ala Leu
165 170 175

Gly Cys Thr Ala Gly Gly Arg Pro Val Gly Ala Phe Gly Asn Ala Glu
180 185 190

Val Phe Ser Phe His Ala Thr Lys Ala Val Thr Ser Phe Glu Gly Gly
195 200 205

Ala Ile Val Thr Asp Asp Gly Leu Leu Ala Asp Arg Ile Arg Ala Met
210 215 220

His Asn Phe Gly Ile Ala Pro Asp Lys Leu Val Thr Asp Val Gly Thr
225 230 235 240

Asn Gly Lys Met Ser Glu Cys Ala Ala Ala Met Gly Leu Thr Ser Leu
245 250 255

Asp Ala Phe Ala Glu Thr Arg Val His Asn Arg Leu Asn His Ala Leu
260 265 270

Tyr Ser Asp Glu Leu Arg Asp Val Arg Gly Ile Ser Val His Ala Phe
275 280 285

Asp Pro Gly Glu Gln Asn Asn Tyr Gln Tyr Val Ile Ile Ser Val Asp
290 295 300

Ser Ala Ala Thr Gly Ile Asp Arg Asp Gln Leu Gln Ala Ile Leu Arg
305 310 315 320

Ala Glu Lys Val Val Ala Gln Pro Tyr Phe Ser Pro Gly Cys His Gln
325 330 335

Met Gln Pro Tyr Arg Thr Glu Pro Pro Leu Arg Leu Glu Asn Thr Glu
340 345 350

Gln Leu Ser Asp Arg Val Leu Ala Leu Pro Thr Gly Pro Ala Val Ser
355 360 365

Ser Glu Asp Ile Arg Arg Val Cys Asp Ile Ile Arg Leu Ala Ala Thr
370 375 380

Ser Gly Glu Leu Ile Asn Ala Gln Trp Asp Gln Arg Thr Arg Asn Gly
385 390 395 400

Ser

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6093 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Streptomyces antibioticus

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 184..1386
- (D) OTHER INFORMATION: /gene= "oleP1"

(ix) FEATURE:

GCATGCCCCG	CTTTCCTCCC	CCTCTCCGAA	CGCATCGACG	ACCCGATCCC	CCTCAGGGAC		60
CGGTGAAGGA	GCGTGTTGCA	CTCATGCAGG	ACATGCAAGG	CGTACAGCCC	GAACCAGCCA		120
GTGTCGAACA	CGCGGCGGAC	GCAGCTCGAA	CAGAGCGAAC	GGCGCACGGA	AGCCGCCCCAG		180
GAG ATG GAG GAC AGC GAA CTG GGG CGC CGC CTG CAG ATG CTC CGC GGC							228
Met Glu Asp Ser Glu Leu Gly Arg Arg Leu Gln Met Leu Arg Gly							
1		5		10		15	
ATG CAG TGG GTC TTC GGC GCC AAC GGC GAT CCG TAC GCC CGG CTG CTG							276
Met Gln Trp Val Phe Gly Ala Asn Gly Asp Pro Tyr Ala Arg Leu Leu							
		20		25		30	
TGT GGC ATG GAG GAT GAC CCG TCA CCT TTC TAC GAC GCG ATA CGG ACC							324
Cys Gly Met Glu Asp Asp Pro Ser Pro Phe Tyr Asp Ala Ile Arg Thr							
		35		40		45	

CTG GGC GAG CTG CAC CGG AGC AGG ACC GGA GCC TGG GTC ACC GCC GAC	372
Leu Gly Glu Leu His Arg Ser Arg Thr Gly Ala Trp Val Thr Ala Asp	
50 55 60	
CCC GGG CTC GGG GGC CGC ATC CTC GCC GAC CGG AAG GCT CGG TGC CCG	420
Pro Gly Leu Gly Gly Arg Ile Leu Ala Asp Arg Lys Ala Arg Cys Pro	
65 70 75	
GAA GGC TCG TGG CCG GTG CGG GCG AAG ACC GAC GGG CTG GAG CAG TAC	468
Glu Gly Ser Trp Pro Val Arg Ala Lys Thr Asp Gly Leu Glu Gln Tyr	
80 85 90 95	
GTG CTG CCC GGG CAC CAG GCG TTC CTG CGG CTG GAG CGC GAG GAG GCC	516
Val Leu Pro Gly His Gln Ala Phe Leu Arg Leu Glu Arg Glu Glu Ala	
100 105 110	
GAG CGA CTG CGG GAG GTC GCG GCG CCG GTG CTG GGG GCC GCG GCG GTC	564
Glu Arg Leu Arg Glu Val Ala Ala Pro Val Leu Gly Ala Ala Ala Val	
115 120 125	
GAC GCG TGG CGC CCG CTG ATC GAC GAG GTC TGC GCG GGG CTC GCG AAG	612
Asp Ala Trp Arg Pro Leu Ile Asp Glu Val Cys Ala Gly Leu Ala Lys	
130 135 140	
GGG CTG CCG GAC ACG TTC GAC CTG GTC GAG GAG TAC GCG GGG CTG GTG	660
Gly Leu Pro Asp Thr Phe Asp Leu Val Glu Glu Tyr Ala Gly Leu Val	
145 150 155	
CCG GTC GAG GTG CTG GCG CGG ATC TGG GGC GTC CCG GAG GAG GAC CGC	708
Pro Val Glu Val Leu Ala Arg Ile Trp Gly Val Pro Glu Glu Asp Arg	
160 165 170 175	
GCC CGG TTC GGG CGT GAC TGC CGG GCG CTC GCT CCC GCG CTG GAC AGC	756
Ala Arg Phe Gly Arg Asp Cys Arg Ala Leu Ala Pro Ala Leu Asp Ser	
180 185 190	
CTC CTG TGT CCC CAG CAG TTG GCG CTG AGC AAG GAC ATG GCG TCC GCC	804
Leu Leu Cys Pro Gln Gln Leu Ala Leu Ser Lys Asp Met Ala Ser Ala	
195 200 205	
CTG GAG GAC CTG CGT CTC CTC TTC GAC GGC CTC GAC GCG ACG CCG CGC	852
Leu Glu Asp Leu Arg Leu Leu Phe Asp Gly Leu Asp Ala Thr Pro Arg	

210	215	220	
CTC GCC GGC CCC GCC GAC GGT GAC GGA ACG GCC GTG GCC ATG CTC ACC			900
Leu Ala Gly Pro Ala Asp Gly Asp Gly Thr Ala Val Ala Met Leu Thr			
225	230	235	
GTT CTG CTC TGC ACG GAG CCG GTG ACC ACG GCG ATC GGG AAC ACC GTG			948
Val Leu Leu Cys Thr Glu Pro Val Thr Thr Ala Ile Gly Asn Thr Val			
240	245	250	255
CTC GGG CTC CTT CCC GGG CAG TGG CCC GTG CCC TGC ACC GGC CGG GTG			996
Leu Gly Leu Leu Pro Gly Gln Trp Pro Val Pro Cys Thr Gly Arg Val			
260	265	270	
GCT GCC GGG CAG GTT GCC GGG CAG GCG CTG CAC CGG GCG GTG TCG TAC			1044
Ala Ala Gly Gln Val Ala Gly Gln Ala Leu His Arg Ala Val Ser Tyr			
275	280	285	
CGT ATC GCG ACG CGG TTC GCC CGG GAG GAC CTG GAG TTG GCG GGC TGC			1092
Arg Ile Ala Thr Arg Phe Ala Arg Glu Asp Leu Glu Leu Ala Gly Cys			
290	295	300	
GAG GTC AAG TCC GGT GAC GAG GTG GTG GTC CTG GCC GGA GCG ATC GGC			1140
Glu Val Lys Ser Gly Asp Glu Val Val Val Leu Ala Gly Ala Ile Gly			
305	310	315	
CGG AAC GGA CCG TCC GCA GCC GCC CCG CCT GCC CCA CCG GGC CCA GCG			1188
Arg Asn Gly Pro Ser Ala Ala Ala Pro Pro Ala Pro Pro Gly Pro Ala			
320	325	330	335
GCC CCG CCC GCC CCG TCG GTC TTC GGT GCC GCC GCC TTC GAG AAC GCG			1236
Ala Pro Pro Ala Pro Ser Val Phe Gly Ala Ala Ala Phe Glu Asn Ala			
340	345	350	
CTG GCC GAA CCC CTC GTC CGG GCT GTG ACG GGA GCG GCC CTC CAG GCC			1284
Leu Ala Glu Pro Leu Val Arg Ala Val Thr Gly Ala Ala Leu Gln Ala			
355	360	365	
CTC GCG GAG GGG CCC CCC CGG CTG ACG GCG GCG GGA CCC GTC GTA CGA			1332
Leu Ala Glu Gly Pro Pro Arg Leu Thr Ala Ala Gly Pro Val Val Arg			
370	375	380	
CGG CGG CGT TCC CCT GTC GTC GGC GGG CTG CAC CGG GCT CCG GTG GCC			1380

Arg Arg Arg Ser Pro Val Val Gly Gly Leu His Arg Ala Pro Val Ala	
385 390 395	
GCC GCA TGAGCATCGC GTCGAACGGC GCGCGCTCGG CCCCCGCCG GCCCCTGCGC	1436
Ala Ala	
400	
GTG ATG ATG ACC ACC TTC GCG GCC AAC ACG CAC TTC CAG CCG CTG GTT	1484
Met Met Met Thr Thr Phe Ala Ala Asn Thr His Phe Gln Pro Leu Val	
1 5 10 15	
CCC CTG GCC TGG GCA CTG CGG ACA GCC GGG CAC GAG GTG CGC GTG GTG	1532
Pro Leu Ala Trp Ala Leu Arg Thr Ala Gly His Glu Val Arg Val Val	
20 25 30	
AGC CAG CCC TCG CTG AGC GAC GTG GTG ACG CAG GCG GGG CTC ACC TCG	1580
Ser Gln Pro Ser Leu Ser Asp Val Val Thr Gln Ala Gly Leu Thr Ser	
35 40 45	
GTC CCG GTG GGC ACC GAG GCT CCG GTC GAG CAG TTC GCG GCG ACC TGG	1628
Val Pro Val Gly Thr Glu Ala Pro Val Glu Gln Phe Ala Ala Thr Trp	
50 55 60	
GGC GAC GAT GCC TAC ATC GGC GTC AAC AGC ATC GAC TTC ACC GGC AAC	1676
Gly Asp Asp Ala Tyr Ile Gly Val Asn Ser Ile Asp Phe Thr Gly Asn	
65 70 75 80	
GAC CCC GGC CTG TGG ACG TGG CCG TAC CTC CTG GGC ATG GAG ACC ATG	1724
Asp Pro Gly Leu Trp Thr Trp Pro Tyr Leu Leu Gly Met Glu Thr Met	
85 90 95	
CTG GTG CCG GCC TTC TAC GAG TTG CTG AAC AAC GAG TCC TTC GTG GAC	1772
Leu Val Pro Ala Phe Tyr Glu Leu Leu Asn Asn Glu Ser Phe Val Asp	
100 105 110	
GGC GTA GTC GAG TTC GCC CGT GAC TGG CCG CCC GAC CTG GTG ATC TGG	1820
Gly Val Val Glu Phe Ala Arg Asp Trp Arg Pro Asp Leu Val Ile Trp	
115 120 125	
GAG CCG CTG ACG TTC GCC GGC GCG GTG GCG GCG CGC GTC ACC GGC GCG	1868
Glu Pro Leu Thr Phe Ala Gly Ala Val Ala Ala Arg Val Thr Gly Ala	
130 135 140	

GCC	CAC	GCC	CGG	CTG	CCG	TGG	GGG	CAG	GAG	ATC	ACC	CTG	CGC	GGG	CGG	1916
Ala	His	Ala	Arg	Leu	Pro	Trp	Gly	Gln	Glu	Ile	Thr	Leu	Arg	Gly	Arg	
145					150					155					160	
CAG	GCG	TTC	CTC	GCC	GAG	CGT	GCC	CTG	CAA	CCG	TTC	GAG	CAC	CGG	GAG	1964
Gln	Ala	Phe	Leu	Ala	Glu	Arg	Ala	Leu	Gln	Pro	Phe	Glu	His	Arg	Glu	
				165					170					175		
GAT	CCC	ACG	GCC	GAG	TGG	CTG	GGC	CGC	ATG	CTC	GAC	CGG	TAC	GGC	TGC	2012
Asp	Pro	Thr	Ala	Glu	Trp	Leu	Gly	Arg	Met	Leu	Asp	Arg	Tyr	Gly	Cys	
			180					185					190			
TCG	TTC	GAC	GAG	GAG	ATG	GTC	ACC	GGG	CAG	TGG	ACC	ATC	GAC	ACG	CTG	2060
Ser	Phe	Asp	Glu	Glu	Met	Val	Thr	Gly	Gln	Trp	Thr	Ile	Asp	Thr	Leu	
		195						200					205			
CCG	CGC	AGC	ATG	CGG	CTG	GAG	CTG	TCC	GAG	GAG	CTG	CGC	ACC	CTG	GAC	2108
Pro	Arg	Ser	Met	Arg	Leu	Glu	Leu	Ser	Glu	Glu	Leu	Arg	Thr	Leu	Asp	
		210					215						220			
ATG	CGG	TAC	GTG	CCG	TAC	AAC	GGA	CCG	GCG	GTC	GTA	CCC	CCC	TGG	GTG	2156
Met	Arg	Tyr	Val	Pro	Tyr	Asn	Gly	Pro	Ala	Val	Val	Pro	Pro	Trp	Val	
225					230					235					240	
TGG	GAA	CCG	TGC	GAG	CGG	CCC	CGG	GTC	TGT	CTG	ACG	ATC	GGC	ACC	TCC	2204
Trp	Glu	Pro	Cys	Glu	Arg	Pro	Arg	Val	Cys	Leu	Thr	Ile	Gly	Thr	Ser	
				245					250					255		
CAG	CGT	GAC	TCC	GGC	CGG	GAC	CAT	GTC	CCC	CTC	GAC	CAC	CTG	CTC	GAC	2252
Gln	Arg	Asp	Ser	Gly	Arg	Asp	His	Val	Pro	Leu	Asp	His	Leu	Leu	Asp	
			260					265					270			
TCC	CTC	GCC	GAC	GTG	GAC	GCG	GAG	ATC	GTG	GCC	ACG	CTC	GAC	ACC	ACC	2300
Ser	Leu	Ala	Asp	Val	Asp	Ala	Glu	Ile	Val	Ala	Thr	Leu	Asp	Thr	Thr	
		275						280					285			
CAG	CAG	GAG	CGC	CTG	CGG	GGC	GCG	GCC	CCC	GGC	AAC	GTC	CGG	CTG	GTG	2348
Gln	Gln	Glu	Arg	Leu	Arg	Gly	Ala	Ala	Pro	Gly	Asn	Val	Arg	Leu	Val	
		290					295				300					
GAC	TTC	GTC	CCG	CTG	CAC	GCG	CTG	ATG	CCG	ACC	TGC	TCG	GCG	ATC	GTG	2396
Asp	Phe	Val	Pro	Leu	His	Ala	Leu	Met	Pro	Thr	Cys	Ser	Ala	Ile	Val	
305					310					315					320	

CAC CAC GGT GGT CCG GGC ACG TGG TCG ACG GCG GCG CTC CAC GGC GTC	2444
His His Gly Gly Pro Gly Thr Trp Ser Thr Ala Ala Leu His Gly Val	
325 330 335	
CCG CAG ATC ATC CTG GAC ACC TCG TGG GAC ACA CCG GTG CGG GCG CAG	2492
Pro Gln Ile Ile Leu Asp Thr Ser Trp Asp Thr Pro Val Arg Ala Gln	
340 345 350	
CGC ATG CAG CAA CTC GGG GCG GGC CTG TCG ATG CCG GTG GGG GAA CTG	2540
Arg Met Gln Gln Leu Gly Ala Gly Leu Ser Met Pro Val Gly Glu Leu	
355 360 365	
GGC GTC GAG GCG CTG CGG GAC CGG GTC CTG CGG CTG CTG GGG GAG CCG	2588
Gly Val Glu Ala Leu Arg Asp Arg Val Leu Arg Leu Gly Glu Pro	
370 375 380	
GAG TTC CGC GCG GGC GCC GAG CGG ATC CGG GCC GAG ATG CTC GCG ATG	2636
Glu Phe Arg Ala Gly Ala Glu Arg Ile Arg Ala Glu Met Leu Ala Met	
385 390 395 400	
CCC GCC CCC GGT GAC GTC GTA CCG GAC CTG GAA CGA CTC ACC GCG GAG	2684
Pro Ala Pro Gly Asp Val Val Pro Asp Leu Glu Arg Leu Thr Ala Glu	
405 410 415	
CAT GCC ACC GGC GCG ATG GCG GGA AGG CGG TGAGACG ATG CGC GTA CTG	2733
His Ala Thr Gly Ala Met Ala Gly Arg Arg Met Arg Val Leu	
420 425 1	
CTG ACC TGC TTC GCC AAC GAC ACC CAC TTC CAC GGG CTG GTG CCG CTG	2781
Leu Thr Cys Phe Ala Asn Asp Thr His Phe His Gly Leu Val Pro Leu	
5 10 15 20	
GCG TGG GCG CTG CGG GCC GCC GGG CAC GAA GTC CGC GTG GCC AGT CAG	2829
Ala Trp Ala Leu Arg Ala Ala Gly His Glu Val Arg Val Ala Ser Gln	
25 30 35	
CCC GCC CTG TCC GAC ACG ATC ACC CAA GCG GGA CTG ACC GCG GTG CCC	2877
Pro Ala Leu Ser Asp Thr Ile Thr Gln Ala Gly Leu Thr Ala Val Pro	
40 45 50	
GTG GGC CGG GAC ACC GCC TTC CTG GAG CTG ATG GGG GAG ATC GGC GCG	2925
Val Gly Arg Asp Thr Ala Phe Leu Glu Leu Met Gly Glu Ile Gly Ala	

55	60	65	
GAC GTC CAG AAG TAC TCC ACC GGC ATC GAC CTG GGC GTC CGC GCG GAG			2973
Asp Val Gln Lys Tyr Ser Thr Gly Ile Asp Leu Gly Val Arg Ala Glu			
70	75	80	
CTG ACG AGC TGG GAG TAC CTG CTC GGC ATG CAC ACG ACC CTG GTG CCC			3021
Leu Thr Ser Trp Glu Tyr Leu Leu Gly Met His Thr Thr Leu Val Pro			
85	90	95	100
ACG TTC TAC TCG CTG GTC AAC GAC GAG CCG TTC GTC GAC GGG CTC GTC			3069
Thr Phe Tyr Ser Leu Val Asn Asp Glu Pro Phe Val Asp Gly Leu Val			
105	110	115	
GCG CTG ACC CGG GCC TGG CGG CCC GAC CTC ATC CTG TGG GAG CAC TTC			3117
Ala Leu Thr Arg Ala Trp Arg Pro Asp Leu Ile Leu Trp Glu His Phe			
120	125	130	
AGC TTC GCC GGG GCG TTG GCG GCG CGG GCC ACC GGC ACG CCC CAC GCC			3165
Ser Phe Ala Gly Ala Leu Ala Ala Arg Ala Thr Gly Thr Pro His Ala			
135	140	145	
CGC GTG CTG TGG GGG TCG GAC CTC ATC GTC CGG TTC CGC CGG GAC TTC			3213
Arg Val Leu Trp Gly Ser Asp Leu Ile Val Arg Phe Arg Arg Asp Phe			
150	155	160	
CTC GCG GAG CGG GCG AAC CGG CCC GCC GAG CAC CGC GAG GAC CCC ATG			3261
Leu Ala Glu Arg Ala Asn Arg Pro Ala Glu His Arg Glu Asp Pro Met			
165	170	175	180
GCG GAG TGG CTG GGC TGG GCG GCC GAA CGG CTG GGC TCC ACC TTC GAC			3309
Ala Glu Trp Leu Gly Trp Ala Ala Glu Arg Leu Gly Ser Thr Phe Asp			
185	190	195	
GAG GAG CTG GTG ACC GGG CAG TGG ACG ATC GAC CCG CTG CCG CGG AGC			3357
Glu Glu Leu Val Thr Gly Gln Trp Thr Ile Asp Pro Leu Pro Arg Ser			
200	205	210	
ATG CGG CTG CCC ACC GGG ACG ACG ACG GTG CCG ATG CGG TAC GTG CCG			3405
Met Arg Leu Pro Thr Gly Thr Thr Thr Val Pro Met Arg Tyr Val Pro			
215	220	225	
TAC AAC GGG CGG GCC GTG GTC CCC GCA TGG GTC CGG CAG CGT GCG CGG			3453

Tyr Asn Gly Arg Ala Val Val Pro Ala Trp Val Arg Gln Arg Ala Arg	
230 235 240	
CGG CCC CGG ATC TGC CTG ACG CTC GGT GTG TCG GCC CGG CAG ACC CTG	3501
Arg Pro Arg Ile Cys Leu Thr Leu Gly Val Ser Ala Arg Gln Thr Leu	
245 250 255 260	
GGC GAC GGC GTG TCG CTG GCG GAG GTG CTG GCC GCG CTG GGC GAC GTG	3549
Gly Asp Gly Val Ser Leu Ala Glu Val Leu Ala Ala Leu Gly Asp Val	
265 270 275	
GAC GCG GAG ATC GTG GCC ACG CTG GAC GCC TCC CAG CGC AAG CTC CTG	3597
Asp Ala Glu Ile Val Ala Thr Leu Asp Ala Ser Gln Arg Lys Leu Leu	
280 285 290	
GGG CCG GTG CCG GAC AAC GTC CGG CTG GTG GAC TTC GTG CCC CTG CAC	3645
Gly Pro Val Pro Asp Asn Val Arg Leu Val Asp Phe Val Pro Leu His	
295 300 305	
GCC CTG ATG CCG ACC TGT TCG GCG ATC GTG CAC CAC GGC GGC GCC GGT	3693
Ala Leu Met Pro Thr Cys Ser Ala Ile Val His His Gly Gly Ala Gly	
310 315 320	
ACC TGG CTG ACG GCC GCC GTC CAC GGC GTC CCG CAG ATC GTC CTC GGT	3741
Thr Trp Leu Thr Ala Ala Val His Gly Val Pro Gln Ile Val Leu Gly	
325 330 335 340	
GAC CTC TGG GAC AAC CTG CTG CGC GCC CGG CAG ACA CAG GCC GCG GGC	3789
Asp Leu Trp Asp Asn Leu Leu Arg Ala Arg Gln Thr Gln Ala Ala Gly	
345 350 355	
GCG GGC CTG TTC ATC CAT CCG TCC GAG GTC ACC GCG GCC GGG CTC GGT	3837
Ala Gly Leu Phe Ile His Pro Ser Glu Val Thr Ala Ala Gly Leu Gly	
360 365 370	
GAG GGC GTG CGC CGG GTG CTG ACG GAC CCT TCC ATC CGG GCC GCC GCA	3885
Glu Gly Val Arg Arg Val Leu Thr Asp Pro Ser Ile Arg Ala Ala Ala	
375 380 385	
CAG CGC GTC CGG GAC GAG ATG AAT GCA GAG CCG ACG CCG GGC GAG GTC	3933
Gln Arg Val Arg Asp Glu Met Asn Ala Glu Pro Thr Pro Gly Glu Val	
390 395 400	

GTC ACG GTG CTG GAG CGG CTC GCC GCG AGC GGC GGA CGC GGA CGA GGA 3981
 Val Thr Val Leu Glu Arg Leu Ala Ala Ser Gly Gly Arg Gly Arg Gly
 405 410 415 420

GGC GGG AAC CAT GCG GGC TGACACGGAG CCGACCACCG GGTACGAGGA 4029
 Gly Gly Asn His Ala Gly
 425

CGAGTTCGCC GAGATCTACG ACGCCGTGTA CCGGGGCCGG GGCAAGGACT ACGCCGGCGA 4089

GGCGAAGGAC GTGGCGGACC TCGTGCGCGA CCGGGTGCCG GACGCGTCCT CCCTCCTGGA 4149

CGTGGCCTGC GGCACGGGCG CGCACCTGCG GCACTTCGCC ACGTCTTCG ACGACGCCCG 4209

CCGTCTCGAA CTGTCCGCGA GCATGCTGGA CATCGCCCGC TCCCGCATGC CGGGCGTGCC 4269

GCTGCACCAA GGGGACATGC GATCCTTCGA CCTGGGGCCA CGCGTCTCCG CGGTACCTG 4329

CATGTTGAGC TCCGTCGGCC ACCTGGCCAC CACCGCCGAA CTCGACGCGA CGCTGCGGTG 4389

CTTGCCCCGG CACACCCGGC CCGGCGGCGT GGCCGTCATC GAACCGTGGT GGTTCCCGGA 4449

GACCTTCACC GACGGCTACG TGGCGGGTGA CATCGTACGC GTCGACGGCC GGACCATCTC 4509

CCGGGTGTCC CACTCGGTAC GGGACGGCGG CGCCACCCGC ATGGAGATCC ACTACGTGAT 4569

CGCCGACGCC GAGCACGGTC CCCGGCACCT GGTGAGCAC CACCGCATCA CGCTGTTCCT 4629

GCGGCATGCG TACACGGCCG CGTACGAGAA GGGGGGCTAC ACCGTCGAGT ACCTCGACGG 4689

CGGGCCCTCG GGCCGGGGGC TGTTCGTCGG CACCCGGACG TGAACCCGCC CGCGCACCGC 4749

CCGATCACCC TGCTCAACGC CGTTCACACG GATCACCGGA CCACGCGAAG GACCTTTCAC 4809

ATG TCG TAC GAC GAC CAC GCG GTG CTG GAA GCG ATA CTG CGG TGC GCC 4857
 Met Ser Tyr Asp Asp His Ala Val Leu Glu Ala Ile Leu Arg Cys Ala
 1 5 10 15

GGA GGT GAC GAG CGC TTC CTG CTG AAC ACC GTC GAG GAA TGG GGA GCC 4905
 Gly Gly Asp Glu Arg Phe Leu Leu Asn Thr Val Glu Glu Trp Gly Ala
 20 25 30

GCC GAG ATC ACC GCG GCG CTC GTG GAC GAG TTG CTG TTC CGC TGC GAG 4953

Ala Glu Ile Thr Ala Ala Leu Val Asp Glu Leu Leu Phe Arg Cys Glu	
35 40 45	
ATC CCG CAG GTG GGC GGT GAG GCG TTC ATC GGC CTG GAC GTC CTG CAC	5001
Ile Pro Gln Val Gly Gly Glu Ala Phe Ile Gly Leu Asp Val Leu His	
50 55 60	
GGC GCC GAC CGG ATC AGC CAT GTG CTG CAG GTG ACG GAC GGC AAG CCG	5049
Gly Ala Asp Arg Ile Ser His Val Leu Gln Val Thr Asp Gly Lys Pro	
65 70 75 80	
GTC ACG TCG GCG GAA CCG GCC GGC CAG GAA CTG GGC GGC CGT ACC TGG	5097
Val Thr Ser Ala Glu Pro Ala Gly Gln Glu Leu Gly Gly Arg Thr Trp	
85 90 95	
AGT TCA CGC TCA GCG ACC CTC CTG CGG GAG CTG TTC GGC CCG CCG TCC	5145
Ser Ser Arg Ser Ala Thr Leu Leu Arg Glu Leu Phe Gly Pro Pro Ser	
100 105 110	
GGC CGC ACC GCG GGG GGC TTC GGC GTC TCC TTC CTG CCC GAC CTG CGC	5193
Gly Arg Thr Ala Gly Gly Phe Gly Val Ser Phe Leu Pro Asp Leu Arg	
115 120 125	
GGC CCG CGG ACC ATG GAG GGC GCG GCC CTG GCC GCC CGC GCC ACC AAC	5241
Gly Pro Arg Thr Met Glu Gly Ala Ala Leu Ala Ala Arg Ala Thr Asn	
130 135 140	
GTG GTG CTG CAC GCG ACG ACC AAC GAG ACG CCC CCA CTG GAC CGG CTG	5289
Val Val Leu His Ala Thr Thr Asn Glu Thr Pro Pro Leu Asp Arg Leu	
145 150 155 160	
GCC CTG CGC TAC GAG TCC GAC AAG TGG GGC GGC GTC CAC TGG TTC ACC	5337
Ala Leu Arg Tyr Glu Ser Asp Lys Trp Gly Gly Val His Trp Phe Thr	
165 170 175	
GGC CAC TAC GAC CGG CAC CTG CGG GCC GTG CGC GAC CAG GCG GTG CGG	5385
Gly His Tyr Asp Arg His Leu Arg Ala Val Arg Asp Gln Ala Val Arg	
180 185 190	
ATC CTG GAG ATC GGC ATC GGC GGC TAC GAC GAC CTG CTG CCG AGC GGC	5433
Ile Leu Glu Ile Gly Ile Gly Gly Tyr Asp Asp Leu Leu Pro Ser Gly	
195 200 205	

GCC TCA CTG AAG ATG TGG AAG CGC TAC TTC CCG CGC GGC CTG GTC TTC	5481
Ala Ser Leu Lys Met Trp Lys Arg Tyr Phe Pro Arg Gly Leu Val Phe	
210 215 220	
GGC GTG GAC ATC TTC GAC AGT CGG CGT GCG ACC AGC CGC GTG TCA AGA	5529
Gly Val Asp Ile Phe Asp Ser Arg Arg Ala Thr Ser Arg Val Ser Arg	
225 230 235 240	
CGC TCC GCG GCC CGG CAG GAC GAC CCG GAG TTC ATG CGC CGC GTC GCC	5577
Arg Ser Ala Ala Arg Gln Asp Asp Pro Glu Phe Met Arg Arg Val Ala	
245 250 255	
GAG GAG CAC GGG CCG TTC GAC GTC ATC ATC GAC GAC GGC AGC CAC ATC	5625
Glu Glu His Gly Pro Phe Asp Val Ile Ile Asp Asp Gly Ser His Ile	
260 265 270	
AAC GCA CAC ATG CGG ACG TCG TTC TCG GTG ATG TTC CCC CAC CTG CGC	5673
Asn Ala His Met Arg Thr Ser Phe Ser Val Met Phe Pro His Leu Arg	
275 280 285	
AAC GGC GGC TTC TAC GTC ATC GAG GAC ACC TTC ACC TCC TAC TGG CCC	5721
Asn Gly Gly Phe Tyr Val Ile Glu Asp Thr Phe Thr Ser Tyr Trp Pro	
290 295 300	
GGG TAC GGA GGG CCA TCC GGA GCC CGG TGC CCG TCC GGA ACA ACC GCG	5769
Gly Tyr Gly Gly Pro Ser Gly Ala Arg Cys Pro Ser Gly Thr Thr Ala	
305 310 315 320	
CTG GAG ATG GTC AAG GGA CTG ATC GAC TCG GTG CAC TAC GAG GAG CGG	5817
Leu Glu Met Val Lys Gly Leu Ile Asp Ser Val His Tyr Glu Glu Arg	
325 330 335	
CCG GAC GGC GCG GCC ACG GCC GAC TAC ATC GCC AGG AAC CTC GTC GGG	5865
Pro Asp Gly Ala Ala Thr Ala Asp Tyr Ile Ala Arg Asn Leu Val Gly	
340 345 350	
CTG CAC GCC TAC CAA ACG ACC TCG TCT TCC TCG AGA AGG GCG ATC AAC	5913
Leu His Ala Tyr Gln Thr Thr Ser Ser Ser Ser Arg Arg Ala Ile Asn	
355 360 365	
AAG GAG GGC GGC ATC CCC CAC ACC GTG CCC CGG GAG CCG TTC TGG AAC	5961
Lys Glu Gly Gly Ile Pro His Thr Val Pro Arg Glu Pro Phe Trp Asn	
370 375 380	

GAC AAC TAGCCACGGC CGCAACCAGA GCCGGAAACC GCACCACTGT CCGCGCCACC 6017
 Asp Asn
 385

TCGGAACCAC CTCCAGCAAA GGACACACCG CTGTGACCGA TACGCACACC GGACCGACAC 6077

CGGCCGACGC GGTACC 6093

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 401 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Met Glu Asp Ser Glu Leu Gly Arg Arg Leu Gln Met Leu Arg Gly Met
 1 5 10 15

Gln Trp Val Phe Gly Ala Asn Gly Asp Pro Tyr Ala Arg Leu Leu Cys
 20 25 30

Gly Met Glu Asp Asp Pro Ser Pro Phe Tyr Asp Ala Ile Arg Thr Leu
 35 40 45

Gly Glu Leu His Arg Ser Arg Thr Gly Ala Trp Val Thr Ala Asp Pro
 50 55 60

Gly Leu Gly Gly Arg Ile Leu Ala Asp Arg Lys Ala Arg Cys Pro Glu
 65 70 75 80

Gly Ser Trp Pro Val Arg Ala Lys Thr Asp Gly Leu Glu Gln Tyr Val
 85 90 95

Leu Pro Gly His Gln Ala Phe Leu Arg Leu Glu Arg Glu Glu Ala Glu
 100 105 110

Arg Leu Arg Glu Val Ala Ala Pro Val Leu Gly Ala Ala Ala Val Asp
 115 120 125

Ala Trp Arg Pro Leu Ile Asp Glu Val Cys Ala Gly Leu Ala Lys Gly
 130 135 140

Leu Pro Asp Thr Phe Asp Leu Val Glu Glu Tyr Ala Gly Leu Val Pro
 145 150 155 160

Val Glu Val Leu Ala Arg Ile Trp Gly Val Pro Glu Glu Asp Arg Ala
 165 170 175

Arg Phe Gly Arg Asp Cys Arg Ala Leu Ala Pro Ala Leu Asp Ser Leu
 180 185 190

Leu Cys Pro Gln Gln Leu Ala Leu Ser Lys Asp Met Ala Ser Ala Leu
 195 200 205

Glu Asp Leu Arg Leu Leu Phe Asp Gly Leu Asp Ala Thr Pro Arg Leu
 210 215 220

Ala Gly Pro Ala Asp Gly Asp Gly Thr Ala Val Ala Met Leu Thr Val
 225 230 235 240

Leu Leu Cys Thr Glu Pro Val Thr Thr Ala Ile Gly Asn Thr Val Leu
 245 250 255

Gly Leu Leu Pro Gly Gln Trp Pro Val Pro Cys Thr Gly Arg Val Ala
 260 265 270

Ala Gly Gln Val Ala Gly Gln Ala Leu His Arg Ala Val Ser Tyr Arg
 275 280 285

Ile Ala Thr Arg Phe Ala Arg Glu Asp Leu Glu Leu Ala Gly Cys Glu
 290 295 300

Val Lys Ser Gly Asp Glu Val Val Val Leu Ala Gly Ala Ile Gly Arg
 305 310 315 320

Asn Gly Pro Ser Ala Ala Ala Pro Pro Ala Pro Pro Gly Pro Ala Ala
 325 330 335

Pro Pro Ala Pro Ser Val Phe Gly Ala Ala Ala Phe Glu Asn Ala Leu
 340 345 350

Ala Glu Pro Leu Val Arg Ala Val Thr Gly Ala Ala Leu Gln Ala Leu
 355 360 365

Ala Glu Gly Pro Pro Arg Leu Thr Ala Ala Gly Pro Val Val Arg Arg
 370 375 380

Arg Arg Ser Pro Val Val Gly Gly Leu His Arg Ala Pro Val Ala Ala
 385 390 395 400

Ala

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 426 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

Met Met Met Thr Thr Phe Ala Ala Asn Thr His Phe Gln Pro Leu Val
 1 5 10 15

Pro Leu Ala Trp Ala Leu Arg Thr Ala Gly His Glu Val Arg Val Val
 20 25 30

Ser Gln Pro Ser Leu Ser Asp Val Val Thr Gln Ala Gly Leu Thr Ser
 35 40 45

Val Pro Val Gly Thr Glu Ala Pro Val Glu Gln Phe Ala Ala Thr Trp
 50 55 60

Gly Asp Asp Ala Tyr Ile Gly Val Asn Ser Ile Asp Phe Thr Gly Asn
 65 70 75 80

Asp Pro Gly Leu Trp Thr Trp Pro Tyr Leu Leu Gly Met Glu Thr Met
 85 90 95

Leu Val Pro Ala Phe Tyr Glu Leu Leu Asn Asn Glu Ser Phe Val Asp
 100 105 110

Gly Val Val Glu Phe Ala Arg Asp Trp Arg Pro Asp Leu Val Ile Trp
115 120 125

Glu Pro Leu Thr Phe Ala Gly Ala Val Ala Ala Arg Val Thr Gly Ala
130 135 140

Ala His Ala Arg Leu Pro Trp Gly Gln Glu Ile Thr Leu Arg Gly Arg
145 150 155 160

Gln Ala Phe Leu Ala Glu Arg Ala Leu Gln Pro Phe Glu His Arg Glu
165 170 175

Asp Pro Thr Ala Glu Trp Leu Gly Arg Met Leu Asp Arg Tyr Gly Cys
180 185 190

Ser Phe Asp Glu Glu Met Val Thr Gly Gln Trp Thr Ile Asp Thr Leu
195 200 205

Pro Arg Ser Met Arg Leu Glu Leu Ser Glu Glu Leu Arg Thr Leu Asp
210 215 220

Met Arg Tyr Val Pro Tyr Asn Gly Pro Ala Val Val Pro Pro Trp Val
225 230 235 240

Trp Glu Pro Cys Glu Arg Pro Arg Val Cys Leu Thr Ile Gly Thr Ser
245 250 255

Gln Arg Asp Ser Gly Arg Asp His Val Pro Leu Asp His Leu Leu Asp
260 265 270

Ser Leu Ala Asp Val Asp Ala Glu Ile Val Ala Thr Leu Asp Thr Thr
275 280 285

Gln Gln Glu Arg Leu Arg Gly Ala Ala Pro Gly Asn Val Arg Leu Val
290 295 300

Asp Phe Val Pro Leu His Ala Leu Met Pro Thr Cys Ser Ala Ile Val
305 310 315 320

His His Gly Gly Pro Gly Thr Trp Ser Thr Ala Ala Leu His Gly Val
325 330 335

Pro Gln Ile Ile Leu Asp Thr Ser Trp Asp Thr Pro Val Arg Ala Gln
 340 345 350

Arg Met Gln Gln Leu Gly Ala Gly Leu Ser Met Pro Val Gly Glu Leu
 355 360 365

Gly Val Glu Ala Leu Arg Asp Arg Val Leu Arg Leu Leu Gly Glu Pro
 370 375 380

Glu Phe Arg Ala Gly Ala Glu Arg Ile Arg Ala Glu Met Leu Ala Met
 385 390 395 400

Pro Ala Pro Gly Asp Val Val Pro Asp Leu Glu Arg Leu Thr Ala Glu
 405 410 415

His Ala Thr Gly Ala Met Ala Gly Arg Arg
 420 425

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 426 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Met Arg Val Leu Leu Thr Cys Phe Ala Asn Asp Thr His Phe His Gly
 1 5 10 15

Leu Val Pro Leu Ala Trp Ala Leu Arg Ala Ala Gly His Glu Val Arg
 20 25 30

Val Ala Ser Gln Pro Ala Leu Ser Asp Thr Ile Thr Gln Ala Gly Leu
 35 40 45

Thr Ala Val Pro Val Gly Arg Asp Thr Ala Phe Leu Glu Leu Met Gly
 50 55 60

Glu Ile Gly Ala Asp Val Gln Lys Tyr Ser Thr Gly Ile Asp Leu Gly
 65 70 75 80

Val Arg Ala Glu Leu Thr Ser Trp Glu Tyr Leu Leu Gly Met His Thr			
	85	90	95
Thr Leu Val Pro Thr Phe Tyr Ser Leu Val Asn Asp Glu Pro Phe Val			
	100	105	110
Asp Gly Leu Val Ala Leu Thr Arg Ala Trp Arg Pro Asp Leu Ile Leu			
	115	120	125
Trp Glu His Phe Ser Phe Ala Gly Ala Leu Ala Ala Arg Ala Thr Gly			
	130	135	140
Thr Pro His Ala Arg Val Leu Trp Gly Ser Asp Leu Ile Val Arg Phe			
	145	150	155
Arg Arg Asp Phe Leu Ala Glu Arg Ala Asn Arg Pro Ala Glu His Arg			
	165	170	175
Glu Asp Pro Met Ala Glu Trp Leu Gly Trp Ala Ala Glu Arg Leu Gly			
	180	185	190
Ser Thr Phe Asp Glu Glu Leu Val Thr Gly Gln Trp Thr Ile Asp Pro			
	195	200	205
Leu Pro Arg Ser Met Arg Leu Pro Thr Gly Thr Thr Thr Val Pro Met			
	210	215	220
Arg Tyr Val Pro Tyr Asn Gly Arg Ala Val Val Pro Ala Trp Val Arg			
	225	230	235
Gln Arg Ala Arg Arg Pro Arg Ile Cys Leu Thr Leu Gly Val Ser Ala			
	245	250	255
Arg Gln Thr Leu Gly Asp Gly Val Ser Leu Ala Glu Val Leu Ala Ala			
	260	265	270
Leu Gly Asp Val Asp Ala Glu Ile Val Ala Thr Leu Asp Ala Ser Gln			
	275	280	285
Arg Lys Leu Leu Gly Pro Val Pro Asp Asn Val Arg Leu Val Asp Phe			
	290	295	300

Val Pro Leu His Ala Leu Met Pro Thr Cys Ser Ala Ile Val His His
305 310 315 320

Gly Gly Ala Gly Thr Trp Leu Thr Ala Ala Val His Gly Val Pro Gln
325 330 335

Ile Val Leu Gly Asp Leu Trp Asp Asn Leu Leu Arg Ala Arg Gln Thr
340 345 350

Gln Ala Ala Gly Ala Gly Leu Phe Ile His Pro Ser Glu Val Thr Ala
355 360 365

Ala Gly Leu Gly Glu Gly Val Arg Arg Val Leu Thr Asp Pro Ser Ile
370 375 380

Arg Ala Ala Ala Gln Arg Val Arg Asp Glu Met Asn Ala Glu Pro Thr
385 390 395 400

Pro Gly Glu Val Val Thr Val Leu Glu Arg Leu Ala Ala Ser Gly Gly
405 410 415

Arg Gly Arg Gly Gly Gly Asn His Ala Gly
420 425

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 386 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

Met Ser Tyr Asp Asp His Ala Val Leu Glu Ala Ile Leu Arg Cys Ala
1 5 10 15

Gly Gly Asp Glu Arg Phe Leu Leu Asn Thr Val Glu Glu Trp Gly Ala
20 25 30

Ala Glu Ile Thr Ala Ala Leu Val Asp Glu Leu Leu Phe Arg Cys Glu
35 40 45

Ile Pro Gln Val Gly Gly Glu Ala Phe Ile Gly Leu Asp Val Leu His
50 55 60

Gly Ala Asp Arg Ile Ser His Val Leu Gln Val Thr Asp Gly Lys Pro
65 70 75 80

Val Thr Ser Ala Glu Pro Ala Gly Gln Glu Leu Gly Gly Arg Thr Trp
85 90 95

Ser Ser Arg Ser Ala Thr Leu Leu Arg Glu Leu Phe Gly Pro Pro Ser
100 105 110

Gly Arg Thr Ala Gly Gly Phe Gly Val Ser Phe Leu Pro Asp Leu Arg
115 120 125

Gly Pro Arg Thr Met Glu Gly Ala Ala Leu Ala Ala Arg Ala Thr Asn
130 135 140

Val Val Leu His Ala Thr Thr Asn Glu Thr Pro Pro Leu Asp Arg Leu
145 150 155 160

Ala Leu Arg Tyr Glu Ser Asp Lys Trp Gly Gly Val His Trp Phe Thr
165 170 175

Gly His Tyr Asp Arg His Leu Arg Ala Val Arg Asp Gln Ala Val Arg
180 185 190

Ile Leu Glu Ile Gly Ile Gly Gly Tyr Asp Asp Leu Leu Pro Ser Gly
195 200 205

Ala Ser Leu Lys Met Trp Lys Arg Tyr Phe Pro Arg Gly Leu Val Phe
210 215 220

Gly Val Asp Ile Phe Asp Ser Arg Arg Ala Thr Ser Arg Val Ser Arg
225 230 235 240

Arg Ser Ala Ala Arg Gln Asp Asp Pro Glu Phe Met Arg Arg Val Ala
245 250 255

Glu Glu His Gly Pro Phe Asp Val Ile Ile Asp Asp Gly Ser His Ile
260 265 270

Asn Ala His Met Arg Thr Ser Phe Ser Val Met Phe Pro His Leu Arg
275 280 285

Asn Gly Gly Phe Tyr Val Ile Glu Asp Thr Phe Thr Ser Tyr Trp Pro
290 295 300

Gly Tyr Gly Gly Pro Ser Gly Ala Arg Cys Pro Ser Gly Thr Thr Ala
305 310 315 320

Leu Glu Met Val Lys Gly Leu Ile Asp Ser Val His Tyr Glu Glu Arg
325 330 335

Pro Asp Gly Ala Ala Thr Ala Asp Tyr Ile Ala Arg Asn Leu Val Gly
340 345 350

Leu His Ala Tyr Gln Thr Thr Ser Ser Ser Ser Arg Arg Ala Ile Asn
355 360 365

Lys Glu Gly Gly Ile Pro His Thr Val Pro Arg Glu Pro Phe Trp Asn
370 375 380

Asp Asn
385

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 738 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Streptomyces antibioticus

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..738
- (D) OTHER INFORMATION: /gene= "oleM"
/note= "SEQ ID No 15 FROM 3992 TO 4729"

(ix) FEATURE:

(A) NAME/KEY: mat_peptide

(B) LOCATION:1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

ATG CGG GCT GAC ACG GAG CCG ACC ACC GGG TAC GAG GAC GAG TTC GCC	48
Met Arg Ala Asp Thr Glu Pro Thr Thr Gly Tyr Glu Asp Glu Phe Ala	
1 5 10 15	
GAG ATC TAC GAC GCC GTG TAC CGG GGC CGG GGC AAG GAC TAC GCC GGC	96
Glu Ile Tyr Asp Ala Val Tyr Arg Gly Arg Gly Lys Asp Tyr Ala Gly	
20 25 30	
GAG GCG AAG GAC GTG GCG GAC CTC GTG CGC GAC CGG GTG CCG GAC GCG	144
Glu Ala Lys Asp Val Ala Asp Leu Val Arg Asp Arg Val Pro Asp Ala	
35 40 45	
TCC TCC CTC CTG GAC GTG GCC TGC GGC ACG GGC GCG CAC CTG CGG CAC	192
Ser Ser Leu Leu Asp Val Ala Cys Gly Thr Gly Ala His Leu Arg His	
50 55 60	
TTC GCC ACG CTC TTC GAC GAC GCC CGC GGT CTC GAA CTG TCC GCG AGC	240
Phe Ala Thr Leu Phe Asp Asp Ala Arg Gly Leu Glu Leu Ser Ala Ser	
65 70 75 80	
ATG CTG GAC ATC GCC CGC TCC CGC ATG CCG GGC GTG CCG CTG CAC CAA	288
Met Leu Asp Ile Ala Arg Ser Arg Met Pro Gly Val Pro Leu His Gln	
85 90 95	
GGG GAC ATG CGA TCC TTC GAC CTG GGG CCA CGC GTC TCC GCG GTC ACC	336
Gly Asp Met Arg Ser Phe Asp Leu Gly Pro Arg Val Ser Ala Val Thr	
100 105 110	
TGC ATG TTC AGC TCC GTC GGC CAC CTG GCC ACC ACC GCC GAA CTC GAC	384
Cys Met Phe Ser Ser Val Gly His Leu Ala Thr Thr Ala Glu Leu Asp	
115 120 125	
GCG ACG CTG CGG TGC TTC GCC CGG CAC ACC CGG CCC GGC GGC GTG GCC	432
Ala Thr Leu Arg Cys Phe Ala Arg His Thr Arg Pro Gly Gly Val Ala	
130 135 140	

```

GTC ATC GAA CCG TGG TGG TTC CCG GAG ACC TTC ACC GAC GGC TAC GTG      480
Val Ile Glu Pro Trp Trp Phe Pro Glu Thr Phe Thr Asp Gly Tyr Val
145                      150                      155                      160

GCG GGT GAC ATC GTA CGC GTC GAC GGC CGG ACC ATC TCC CGG GTG TCC      528
Ala Gly Asp Ile Val Arg Val Asp Gly Arg Thr Ile Ser Arg Val Ser
                      165                      170                      175

CAC TCG GTA CGG GAC'GGC GGC GCC ACC CGC ATG GAG ATC CAC TAC GTG      576
His Ser Val Arg Asp Gly Gly Ala Thr Arg Met Glu Ile His Tyr Val
                      180                      185                      190

ATC GCC GAC GCC GAG CAC GGT CCC CGG CAC CTG GTC GAG CAC CAC CGC      624
Ile Ala Asp Ala Glu His Gly Pro Arg His Leu Val Glu His His Arg
                      195                      200                      205

ATC ACG CTG TTC CCG CGG CAT GCG TAC ACG GCC GCG TAC GAG AAG GCG      672
Ile Thr Leu Phe Pro Arg His Ala Tyr Thr Ala Ala Tyr Glu Lys Ala
                      210                      215                      220

GGC TAC ACC GTC GAG TAC CTC GAC GGC GGG CCC TCG GGC CGG GGG CTG      720
Gly Tyr Thr Val Glu Tyr Leu Asp Gly Gly Pro Ser Gly Arg Gly Leu
225                      230                      235                      240

TTC GTC GGC ACC CGG ACG      738
Phe Val Gly Thr Arg Thr
                      245

```

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 246 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

```

Met Arg Ala Asp Thr Glu Pro Thr Thr Gly Tyr Glu Asp Glu Phe Ala
  1              5              10              15

Glu Ile Tyr Asp Ala Val Tyr Arg Gly Arg Gly Lys Asp Tyr Ala Gly

```

20	25	30
Glu Ala Lys Asp Val Ala Asp Leu Val Arg Asp Arg Val Pro Asp Ala		
35	40	45
Ser Ser Leu Leu Asp Val Ala Cys Gly Thr Gly Ala His Leu Arg His		
50	55	60
Phe Ala Thr Leu Phe Asp Asp Ala Arg Gly Leu Glu Leu Ser Ala Ser		
65	70	75 80
Met Leu Asp Ile Ala Arg Ser Arg Met Pro Gly Val Pro Leu His Gln		
85	90	95
Gly Asp Met Arg Ser Phe Asp Leu Gly Pro Arg Val Ser Ala Val Thr		
100	105	110
Cys Met Phe Ser Ser Val Gly His Leu Ala Thr Thr Ala Glu Leu Asp		
115	120	125
Ala Thr Leu Arg Cys Phe Ala Arg His Thr Arg Pro Gly Gly Val Ala		
130	135	140
Val Ile Glu Pro Trp Trp Phe Pro Glu Thr Phe Thr Asp Gly Tyr Val		
145	150	155 160
Ala Gly Asp Ile Val Arg Val Asp Gly Arg Thr Ile Ser Arg Val Ser		
165	170	175
His Ser Val Arg Asp Gly Gly Ala Thr Arg Met Glu Ile His Tyr Val		
180	185	190
Ile Ala Asp Ala Glu His Gly Pro Arg His Leu Val Glu His His Arg		
195	200	205
Ile Thr Leu Phe Pro Arg His Ala Tyr Thr Ala Ala Tyr Glu Lys Ala		
210	215	220
Gly Tyr Thr Val Glu Tyr Leu Asp Gly Gly Pro Ser Gly Arg Gly Leu		
225	230	235 240
Phe Val Gly Thr Arg Thr		
245		

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

TCCTCGATGG AGACCTGCC

19

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

GAGACCATGC CCAGGGAGT

19

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

TCTGGGAGCC GCTCACCTT

19

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

GACGAGGCCG AAGAGGTGG

19

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

GCACACCGGA ATGGATGCG

19

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

CCGTCGAGCT CTGAGGTAA

19

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

GCCCCGAGCCG CACGTGCGT

19

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

TGCACGCGCT GCTGCCGACC

20

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

TTGGCGAAGT CGACCAGGTC

20

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

GCCGCTCGGC ACGGTGAACT TCA

23

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

ATGCGCGTCG TCTTCTCCTC CATG

24

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

TCATCGTGGT TCTTCCTTC C

21

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

GGAATTCATG ACCACGACCG ATC

23

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

CGCTCCAGGT GCAATGCCGG GTGCAGGC

28

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

GATCACGCTC TTCGAGCGGC AG

22

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

GAACTCGGTG GAGTCGATGT C

21

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

GTTGTCGATC AAGACCCGCA C

21

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

CATCGTCAAG GAGTTCGACG GT

22

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

TGCGCAGGTC CATGTTACC ACGTT

25

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

GCTACGCCCT GGAGAGCCTG

20

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

GTCGCGGTCG GAGAGCACGA C

21

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

GCCAGCTCGG CGACGTCCAT C

21

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

CGACGAGGTC GTGCATCAG

19

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

AATTGATCAA GGTGAACACG GTCATGCGCA GGATCCTCGA GCGGAAGTCC ATGGGG

56

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 56 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

CCCCATGGAG TTCCGCTCGA GGATCCTGCG CATGACCGTG TTCACCTGA TCAATT

56

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

AACTCGGTGG AGTCGATGTC GTCGCTGCGG AA

32

(2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

CAATATAGGA AGGATCAAGA GGTGAC

27

(2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

TCCGGAGGTG TGCTGTCGGA CGGACTTGTC GGTCGGAAA

39

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

AGGAGCACTA GTGCGGGTAC TGCTGACGTC CTT

33

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

GGGGGATCCC ATATGCGGGT ACTGCTGACG TCCTTCG

37

(2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

GAAAAGATCT GCCGGCGTGG CGGCGCGTGA GTTCCTC

37

(2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

AGCGGCTTGA TCGTGTGGA CCAGTAC

27

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

GGCCTATGTG GACTACGTGT TGAACGT

27

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

AACGCCTCGT CCTGCAGCGG AGACACGAAC A

31

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

TTCGCTCCCC GATGAACACA ACTCGTA

27

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 35 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

GAAGGAGATA TACATATGCG CGTCGTCTTC TCCTC

35

(2) INFORMATION FOR SEQ ID NO: 58:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

CGGGATCCTC ATCGTGGTTC TCTCCTTCCT GC

32

(2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

CGGGTACCAT GCGCGTCGTC TTCTCCTCCA TG

32

(2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

CGGGTACCTC ATCGTGGTTC TCTCCTTCC

29

(2) INFORMATION FOR SEQ ID NO: 61:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Peptide
- (B) LOCATION:1..13
- (D) OTHER INFORMATION:/note= "SEQ ID No 11 FROM 38 TO 50"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

Val Thr Gly Ala Gly Asp Gly Asp Ala Asp Val Gln Ala
1 5 10